INVESTIGATION OF SMAD-TRANSCRIPTION FACTOR INTERACTIONS ON THE *c-Jun* PROMOTER

BY

SEAN HARKIN

A THESIS SUBMITTED TO THE UNIVERSITY OF LONDON
FOR THE DEGREE OF MASTER OF PHILOSOPHY

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

There are a lot of people I would like to thank for their support in carrying out this project. Firstly, my supervisor, Dr. Caroline Hill for giving me the chance to do this work and for providing me with the benefit of her scientific expertise. Also my thesis Committee – Dr. Richard Treisman, Dr. Sally Leevers and Dr. David Ish Horowicz – for their advice and suggestions. Dr. Peter Parker for his professional guidance and for helping me to think it all through. Dr. Mike Howell in the Developmental Signalling Laboratory for sharing his expertise and for patiently answering endless technical questions. The other members of the lab for their scientific help and for providing a fun and friendly working environment. Cancer Research UK scientific services – in particular Cell Services, Antibody Production, Oligonucleotide Synthesis and Light Microscopy – for providing me with dependable support and experimental reagents. My former lecturers, Dr. Maura Grealy, Dr. Maria Tuohy and Dr. Pat Morgan for providing a fresh perspective and for taking the time to give me detailed opinions. My mother and my grandparents, for always putting me first and for giving me everything I need to deal with the challenges of life. Fiona, for being a wee ray of Scottish sunshine and for making life so much fun! My friends – Karen, Stu, Denis, Edel, Jonathan, John, Shane, Philip, Cedric, Sarah, Emilie and everyone else – for their support in difficult times and for making sure that I could always take my mind off it all. And finally, Mary, for her help with the big decisions.

ABSTRACT

Members of the TGF-β superfamily are secreted growth and differentiation factors that regulate diverse events in embryonic development and adult tissue homeostasis. Signalling by mammalian TGF-β involves phosphorylation of Smad2 and Smad3 by TGF-β receptor complexes, association with Smad4, nuclear accumulation of Smad complexes, DNA binding and regulation of target gene expression. Smad complexes rely on associated transcription factors and coregulators for gene regulation and availability of different partners may underlie the cell-type specificity of TGF-β responses. Several Smad-interacting transcription factors have been well characterised in *Xenopus* embryos but, despite numerous studies, the identities and functions of Smad-interacting transcription factors in adult mammalian tissues remain poorly understood. Furthermore, the best-characterised Smad-interacting transcription factors bind Smad2-Smad4 complexes rather than Smad3-Smad4 complexes. Previous work has demonstrated that a TGF-β-inducible complex forming on the 55-basepair *c-Jun* Smad-binding region (SBR) contains a Smad3-Smad4 heterodimer plus two unidentified factors. My objectives were to further characterise this complex *in vitro* and develop a purification strategy for identification of its components by mass spectrometry. In this thesis I show that TGF-β-induced complex formation on the SBR is dependent on Smad4. I have also found that the minimal SBR is 32 basepairs long and contains a $(GCCA)_2$ site to which one of the unidentified factors binds. Similarly, sequential pull-downs confirm the presence of at least one additional factor that is essential for association of the Smad3-Smad4 dimer with the SBR and which only associates with the Smads on DNA. The thesis also describes a two-step affinity purification strategy that I have developed to isolate the components of the complex for identification by mass spectrometry. Finally, I have shown that Smad4 and the corepressor SnoN associate with the SBR in the absence of TGF-β, suggesting the possibility of a repressive complex. These studies extend the understanding of Smad-partner interactions and could lead to identification of transcriptional partners of Smad3-Smad4 complexes in adult cells.

CONTENTS

Contents

Contents

LIST OF FIGURES/TABLES

FIGURES

Chapter 1

Chapter 3

TABLES

Chapter 2

ABBREVIATONS

ActRIIA	Activin receptor type II A
ActRIIB	Activin receptor type II B
ALK	Activin receptor-like kinase
AMH	Anti-Mullerian hormone (same protein as MIS)
AMHRII	Anti-Mullerian hormone receptor type II
AMP	Adenosine monophosphate
AMPK1	AMP-activated protein kinase
AP1	Activating protein 1
ARE	Activin response element
ARF	Activin response factor
ATF	Activating transcription factor
ATP	Adenosine triphosphate
bHLH	Basic domain – helix-loop-helix
BMP	Bone morphogenetic protein
BMPRII	Bone morphogenetic protein receptor type II
BrdU	Bromodeoxyuridine
bZIP	Basic domain – leucine zipper
cAMP	Cyclic adenosine monophosphate
CARM1	Coactivator-associated arginine methyltransferase 1
CBP	CREB-binding protein
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CHRAC	Chromatin accessibility complex
CKI	Cyclin kinase inhibitor
Co-IP	Co-immunoprecipitation
Co-Smad	Common mediator Smad
CpG	Cytosine-phosphate-guanosine
CPM	Counts per minute
CRE	cAMP response element
CREB	CRE-binding protein

CRM1	Chromosome region maintenance 1
CtBP	C-terminal binding protein
C-terminal	Carboxy-terminal
D1-D4	Downstream 1 to downstream 4
DAG	Diacylglycerol
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxyribonucleic acid
DNAP	DNA pull-down
DNase	Deoxyribonuclease
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleoside triphosphate
dTTP	Deoxythymidine triphosphate
DMPD	Dimethylpimelimidate
Dpp	Decapentaplegic
EMSA	Electrophoretic mobility shift assay
EMT	Epithelial-to-mesenchymal transition
ERK	Extracellular signal-regulated kinase
Fast	Forkhead activin signal transducer
FKBP12	FK506-binding protein
FM	Fast motif
Fox	Forkhead box
GADD	Growth arrest- and DNA damage-inducible gene
Gcn5	General control of amino acid synthesis 5
GDF	Growth and differentiation factor
GDP	Guanosine diphosphate
GHRH	Growth hormone releasing hormone
GPCR	G-protein coupled receptor
GS	Glycine-serine motif
GSK	Glycogen synthase kinase
GTF	General transcription factor
GTP	Guanosine triphosphate
H1-H4	Histones H1 to H4

HA	Haemaglutinin
HAT	Histone acetyltransferase
HBV	Hepatitis B virus
HDAC	Histone deacetylase
HGF	Hepatocyte growth factor
HLH	Helix-loop-helix
HMG	High-mobility group
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
HSF	Heat-shock factor
Hsp	Heat-shock protein
HTLV	Human T-cell lymphotropic virus
IFN	Interferon
IP	Immunoprecipitation
IP_3	Inositol triphosphate
IRF	Interferon response factor
I-Smad	Inhibitory Smad
JKAP	JNK pathway-associated phosphatase
JNK	Jun N-terminal kinase
kD	Kilodalton
KID	Kinase-inducible domain
KSHV	Kaposi's sarcoma-associated herpesvirus
Lef	Lymphoid enhancer factor
Mad	Mothers against decapentaplegic
MAPK	Mitogen-activated protein kinase
MDCK	Madin-Darby canine kidney
MEF	Mouse embryonic fibroblast or muscle enhancer factor
MEK	MAPK/ERK kinase
MH1	Mad-homology 1 domain
MH2	Mad-homology 2 domain
MIS	Mullerian inhibiting substance (same protein as AMH)
MMP	Matrix metalloproteinase
MSK	Mitogen- and stress-activated protein kinase 1
Mut	Mutant

N-CoR	Nuclear receptor corepressor
NES	Nuclear export signal
NET1	Neuroepithelial cell transforming factor 1
NF1	Nuclear factor 1
NF-κB	Nuclear factor kappa B
NLS	Nuclear localisation signal
N-terminal	Amino-terminal
Nup	Nucleoporin
ORF	Open reading frame
PAI-1	Plasminogen activator inhibitor 1
PBSA	Phosphate buffered saline A (standard composition)
PBST	Phosphate buffered saline with Tween-20
PCAF	p300/CBP-associated factor
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PDK1	Phosphatidylinositol-dependent kinase 1
PKA	Protein kinase A
PKB	Protein kinase B
PKC	Protein kinase C
PI3K	Phosphatidylinositol-3-kinase
PLC	Phospholipase C
$PI(3,4)P_2$	Phosphatidylinositol-3,4-bisphosphate
$PI(3,4,5)P_3$	Phosphatidylinositol-3,4,5-triphosphate
$PI(4,5)P_2$	Phosphatidylinositol-4,5-bisphosphate
PP1	Protein phosphatase 1
PP2A	Protein phosphatase 2A
PRMT1	Protein arginine methyltransferase 1
Rb	Retinoblastoma protein
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
ROCK	Rho-activated kinase
RNase	Ribonuclease
RSF	Remodelling and spacing factor

R-Smad	Receptor-regulated Smad
RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcriptase PCR
SAD	Smad activation domain
SARA	Smad anchor for receptor activation
SBE	Smad binding element
SBR	Smad binding region
SDS	Sodium dodecyl sulphate
SIP	Smad interacting protein
Ski	Sloan-Kettering Institute avian retrovirus
Sno	Ski-related novel gene
shRNA	Short hairpin RNA
SID	Smad interaction domain
SIM	Smad interaction motif
Sma	Small phenotype
Smad	Sma-Mad homolog
Smurf	Smad ubiquitin regulatory factor
Sp	Specificity protein
SUMO	Small ubiquitin-like modifier
SWI/SNF	Switching/sucrose non fermenting
TAF	TBP-associated factor
TBP	TATA-binding protein
TβRI	TGF-β receptor type I
TβRII	TGF-β receptor type II
TCF	Ternary complex factor
TEMED	N-N-N'-N'-tetramethylenediamine
Tet	Tetracycline
TFE3	Transcription factor E3
TFIIA – TFIIJ	Transcription factors IIA to IIJ
TFTC	TBP-free TAF_{II} complex
TGF-α	Transforming growth factor alpha
TGF-β	Transforming growth factor beta
TGIF	TG-interacting factor

TIEG	TGF-β-inducible early gene
TIMP	Tissue inhibitor of metalloproteinase
TLP	TRAP1-like protein
Tr	Truncated
TRAP1	TNF-α associated protein 1
TRAP220	Thyroid hormone receptor associated protein 220
U1-U4	Upstream 1 to upstream 4
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
WT	Wild-type
Xnr	*Xenopus* Nodal-related

CHAPTER 1
INTRODUCTION

1.1 Biological signalling and the control of gene expression

As an animal grows and develops from a single cell into a mature organism it requires exquisite coordination. Bodily regions must be specified and masses of cells must simultaneously grow, divide, differentiate, migrate or die in order to produce the immensely complex patterning that is characteristic of the fully formed animal (Slack, 2001). In the adult organism too the activities of individual cells must be tightly coordinated for tissue homeostasis and proper physiological function. The biochemical mechanisms by which this coordination is achieved have been the subject of intense investigation (Strewler, 1997).

The central process underlying biological coordination in multicellular organisms is the transmission of signals between cells, either by direct cell-cell contact or through diffusible biochemical messengers. In animals, these messengers usually take the form of proteins or peptides that are released by one cell and bind to the extracellular domain of receptor proteins on the surface of recipient cells (Helmreich, 2001). Such ligand-receptor binding leads to multimerisation of receptors and/or conformational changes that stimulate the activity of the receptors' intracellular domains, allowing them to activate specific cytoplasmic signalling molecules. This initiates a cascade of events that typically culminates in the activation or inhibition of specific transcription factors and the induction or repression of a particular set of target genes. Changes in the expression level of target gene products are then responsible for the cellular response to the signal received (Lee and Goodbourn, 2001).

Extracellular signals can also initiate cellular responses independently of transcription. Cellular responses that operate independently of transcriptional control include effects on gene expression at other levels, effects on the activity of key metabolic enzymes and effects on cytoskeletal dynamics (Helmreich, 2001).

1.1.1 Signal transduction pathways

The physical details of signalling pathways vary greatly. Some, such as growth factor signalling via receptor tyrosine kinases (RTKs) and mitogen-activated protein kinases (MAPKs), operate via a cascade of protein kinase activations that lead to phosphorylation and activation of nuclear transcription factors. Others, such as cytokine signalling via the Jak-Stat pathway, involve phosphorylation of transcription factors at the plasma membrane. These activated transcription factors then accumulate in the nucleus and regulate gene expression (Lee and Goodbourn, 2001). Still other signalling pathways operate by regulated proteolysis. Binding of the receptor Notch by its ligand Delta, for example, leads to proteolytic cleavege of the Notch intracellular domain and accumulation of this fragment in the nucleus, where it binds specific DNA sites in cooperation with other factors and acts to regulate transcription (Kadesch, 2004).

A further signalling mechanism used by protein and peptide hormones, as well as small-molecule neurotransmitters, is activation of trimeric G-proteins via G-protein coupled receptors (GPCRs). These G-proteins consist of α, β and γ subunits. When activated, the α subunit releases its bound GDP molecule and binds GTP in its place. This leads to dissociation of the α subunit and a $\beta\gamma$ subcomplex, both of which activate target proteins by direct physical association (Marinissen and Gutkind, 2001). Similarly, monomeric G proteins such as Ras, Rho, Rac and Cdc42 are activated downstream of RTKs, exchange GDP for GTP and are involved in MAPK pathways and other signalling cascades (Bar-Sagi and Hall, 2000).

Signalling downstream of cell-surface receptors can also involve production of small-molecule second messengers in the cytoplasm or plasma membrane. These include phosphatidylinositol devivatives, Ca^{2+} and cyclic AMP (cAMP) (Helmreich, 2001). Activation of phospatidylinositol-3-kinase (PI3K) at RTKs and GPCRs leads to phosphoryation of membrane-bound inositol phospholipids to produce $PI(3,4)P_2$ or $PI(3,4,5)P_3$. These provide a docking site for specific protein kinases, such as protein kinase B (PKB)/Akt that transmit the signal onwards (Leevers et al., 1999). Phospholipase C (PLC) is also activated via RTKs or GPCRs and cleaves $PI(4,5)P_2$ to

produce inositol triphosphate (IP_3) and diacylglycerol (DAG). Diffusible IP_3 stimulates the release of Ca^{2+} from the endoplasmic reticulum while membrane-bound DAG provides a platform for the activation of protein kinases, including protein kinase C (PKC) (Williams and Katan, 1996). Similarly, adenlyate cyclase is activated by specific trimeric G proteins and catalyses the formation of cAMP from ATP. cAMP then binds and activates specific signalling mediators, including protein kinase A (PKA) (De Cesare et al., 1999).

Finally, peptide hormones and small-molecule neurotransmitters can directly activate ligand-gated ion channels, leading to changes in the transmembrane dynamics of ions such as Na^+, K^+, Ca^{2+} and Cl^-. This is the mechanism by which action potentials are transmitted across neural synapses (Jackson, 1999).

Steroid hormones, vitamin D and nitric oxide signal in a fundamentally different way from protein and peptide messengers. These can pass through the plasma membrane and bind directly to intracellular targets. Steroids and vitamin D bind to nuclear receptors that also contain DNA-binding domains and function directly as transcription factors (McKenna and O'Malley, 2002).

From this brief overview it is clear that diverse mechanisms underlie the cellular response to extracellular signals. However, a common theme in the majority of signalling pathways is activation or inhibition of nuclear transcription factors and up- or down-regulation of transcription initiation at specific target gene promoters. Since transcription initiation is the first step in gene expression, modulating the process at this level can have the greatest quantitative effect on gene product expression and is the primary level at which cellular differentiation and function are regulated (Lee and Goodbourn, 2001).

1.1.2 Transcription factors

Regulatory transcription factors are proteins found in eukaryotic cells that function to control gene expression at the level of RNA synthesis. They contain both a DNA-binding domain and a transcription activation/repression domain. The DNA binding

domain associates directly with specific sites in DNA – providing the basis for target gene selection – while the activation/repression domain controls the expression of the adjacent gene (Kadonaga, 2004). Different activation mechanisms exist but they generally culminate in the recruitment of RNA polymerase (RNAP II in the case of protein-coding genes) and the basal transcription machinery, followed by transcription initiation. Conversely, different repression mechanisms involve inhibition of this process. Regulatory transcription factors may also contain dimerisation domains and additional regions that are required for control by upstream signals (Kadonaga, 2004).

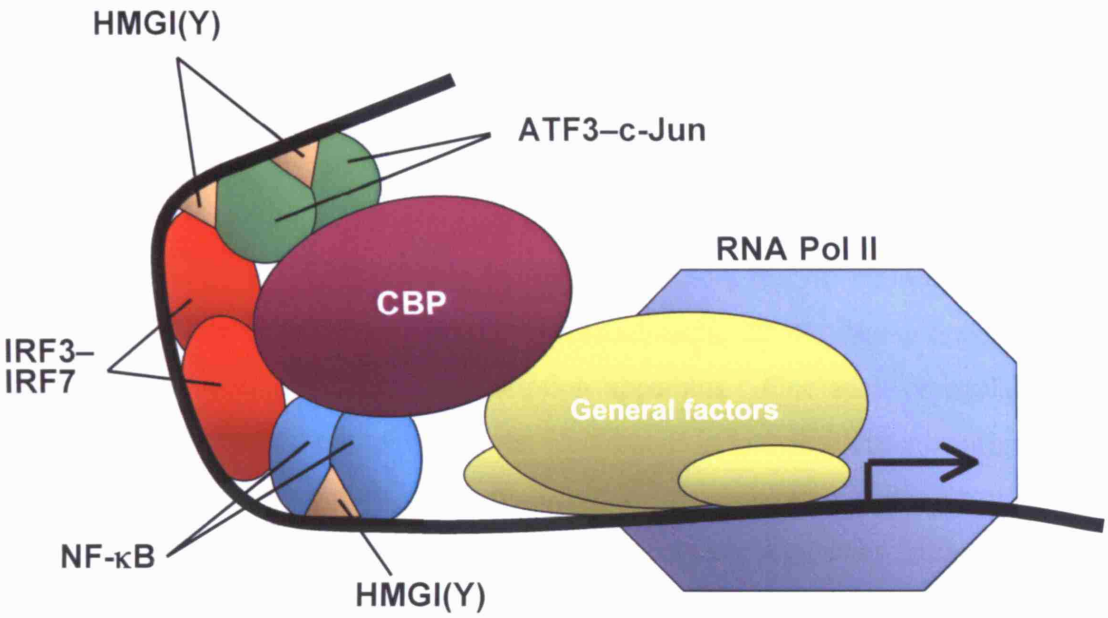
Regulatory transcription factors can be divided into different categories on the basis of the kind of DNA binding domain they contain. The principal categories include helix-turn-helix (HTH) factors, zinc finger proteins, basic domain-leucine zipper (bZIP) factors and basic helix-loop-helix (bHLH) proteins. In the case of bZIP and bHLH factors, a characteristic dimerisation motif is associated with a common basic DNA-binding domain (Pabo and Sauer, 1992). For activation/repression domains, well-defined structural motifs have been more difficult to identify. The first known activation motif was the "acid blob", a single amphiphatic α-helix with one acidic face and one hydrophobic face that is sufficient for transcriptional activation (Giniger and Ptashne, 1987; Ptashne, 1988). Other motifs enriched in glutamine or proline or containing hydrophobic β-sheets have since been identified. In recent years it has become clear that these motifs function as protein-protein interaction sites for the recruitment of transcriptional coactivators. These activate transcription either by recruiting the basal transcription machinery for RNAPII or by converting chromatin from an "inactive" to an "active" state (Ptashne and Gann, 1997). Structural motifs involved in transcription repression are poorly understood but it is thought that they function in a similar fashion by recruiting transcriptional corepressors (Kadonaga, 2004).

Sequence-specific transcription factors can be regulated by post-translational modification in response to upstream signals. The cAMP response factor CREB, for example, is a bZIP factor that induces target gene expression downstream of hormones such as glucagon and GHRH (De Cesare et al., 1999; Mayr and Montminy,

2001). In these pathways, CREB is phosphorlyated by protein kinase A (PKA) on Ser133 in its kinase-inducible domain (KID) (Gonzalez and Montminy, 1989). This allows it to bind the cAMP response element (CRE) of target gene promoters and activate transcription by recruiting the coactivator CBP via the KID and $TAF_{II}130/135$ (a component of the general transcription factor TFIID – see below) via two glutamine-rich regions flanking the KID (Nakajima et al., 1997). Generally, transcription factor activity (and subcellular localisation) can be regulated by phosphorylation, acetylation, ubiquitination, SUMOylation and other modifications (Kadonaga, 2004).

Eukaryotic promoters and enhancers contain multiple transcription factor binding sites in a specific linear arrangement. Association of transcription factors with these sites is cooperative, involving direct protein-protein interactions and DNA bending to create more favourable binding sites for other factors. This allows stable regulatory complexes to form on these regions (Chen, 1999; Courey, 2001). The virus-induced complex that forms on the *IFNβ* enhancer, for example, contains the transcription factors NF-κB, IRF3, IRF7, ATF-2 and c-Jun as well as the architectural factor HMGI(Y), all in a specific physical arrangement. Several of these factors interact with the coactivator CBP, creating a high-affinity "landing-pad" for this protein, which then recruits the basal transcription apparatus (Fig. 1.1) (Thanos and Maniatis, 1995). Such complexes function in a combinatorial fashion and incorporation of different factors gives rise to different levels of transcription. In many cases – including the *IFNβ* enhancer – the absence of a single factor greatly destabilises the complex. This makes gene expression highly dependent on the presence of multiple signals and allows the same signalling cascade to induce different gene sets in different contexts. As a result, every gene in an animal has a unique and tightly regulated pattern of expression at every stage of development and in every physiological state (Chen, 1999; Courey, 2001).

Along with regulatory factors, eukaryotic cells also contain general transcription factors (GTFs). These are components of the basal transcription machinery that may or may not bind directly to DNA but which are all involved in recognition of the core promoter and initiation of transcription. The GTFs associated with RNAPII include



Adapted from Courey, 2001

Fig. 1.1 Transcription factor cooperation. Following viral infection of mammalian cells, various stress signals induce the formation of a multi-protein complex on the *IFNβ* promoter. This consists of the transcription factors NF-κB, IRF3, IRF7, ATF-2 and c-Jun; the coactivator CBP and the architectural factor HMGI(Y). The transcription factors together create a high-affinity "landing pad" for CBP, which then directly recruits the basal transcription apparatus. If any subunit is unavailable a stable complex cannot form, CBP is not recruited and transcription is not induced. This is a good example how promoter activation can be made dependent on multiple signals.

TFIIA to TFIIF, TFIIH and TFIIJ. Several of these are multi-subunit complexes (Latchmann, 2002; Woychik and Hampsey, 2002) and, within these complexes, specific subunits have been shown to have regulatory functions. $TAF_{II}30$/TAF10, for example, is a subunit of TFIID that has been shown to be required for keratinocyte differentiation in mouse embryos but which is dispensable for tissue homeostasis in the adult epidermis (Indra et al., 2005).

1.1.3 Coregulators

The mechanism by which eukaryotic transcription factors control transcription is by recruitment of coactivators and corepressors (Kadonaga, 2004). Some coactivators function by recruiting the basal transcription apparatus. One such coregulator is Mediator, a large protein complex (at least 25 subunits in humans) that was originally identified by biochemical purification of factors essential for transcription activation *in vitro*. It appears to function as a physical bridge between regulatory factors and the basal transcription apparatus. The mediator subunit TRAP220/Med1, for example, seems to target mediator to nuclear hormone receptors (Zhang et al., 2005).

Different forms of the mediator exist and are thought to exert either an activating or repressing influence on the assembly of the basal transcription apparatus depending on their precise subunit composition (Conaway et al., 2005). TRAP220/Med1 has been shown to play a role in transcription activation by many (but not all) forms of mediator (Zhang et al., 2005). Similarly, in yeast, Hrs1/Med3 is a mediator subunit that recruits a repressive form of the complex to specific target promoters (Papamichos-Chronakis et al., 2000).

Other coactivators that function by recruiting the basal transcription apparatus include steroid receptor cocoactivator 1 (SRC1), which interacts with TBP and TFIIB (Takeshita et al., 1996). Likewise, the viral transactivator VP16 interacts directly with $TAF_{II}31$ (Uesugi et al., 1997). In a related manner, the transcription factors Oct1 and Oct2 can directly recruit TFIID to induce target gene transcription (Zwilling et al., 1994).

However, the mechanisms of coactivator function that have received most attention in recent years are covalent modification of histones or DNA and chromatin remodelling or assembly. Many coregulators are enzymes that modify core histones on their N-terminal tails or at other sites. Mammalian coactivators that function in this way include the histone acetyltransferases (HATs) and histone methyltransferases (HMTs), both of which act on lysines and arginines (Iizuka and Smith, 2003). HATs include CBP, p300 (a CBP paralog), Tip60, PCAF and Gcn5. HMTs include CARM1, PRMT1 and Set9. Mammalian histone-modifying corepressors include the histone deacetylases (HDACs), HMTs such as SETDB1 and G9a and the histone kinases AMPK and MSK1. HATs have also been reported to be involved in repression (Peterson and Laniel, 2004). Other coregulators are DNA methytransferases (DNMTs) that silence genes by methylating cytosine bases in the CpG islands that surround certain promoters (Brenner and Fuks, 2006).

Another group of enzymes that function as coactivators are ATP-dependent DNA translocases that remodel chromatin (Flaus and Owen-Hughes, 2004). Mammalian members of this group include Brg1, Brm, Snf2h and Mi-2. Many details of how exactly these proteins function remain unclear, but what is known is that they act both to reposition nucleosomes along the chromatin fibre and to increase the availability of DNA within the histone octamer to proteins from the nucleoplasm (Narlikar et al., 2002).

All of these factors tend to reside in large multiprotein complexes that are recruited to chromatin by transcription factors. Brg1 and Brm are alternative catalytic subunits of the SWI/SNF complex, SNF2h is the catalytic subunit of the CHRAC, RSF and ACF complexes and Mi-2 resides in the NuRD complex (Narlikar et al., 2002; Wang et al., 1996). Similarly, PCAF and Gcn5 (which are paralogs of each other) are subunits of the STAGA, PCAF and TFTC complexes; Tip60 is a subunit of the NuA4 complex and HDAC1 and HDAC2 are subunits of the Sin3 and NuRD complexes (Doyon et al., 2004; Grant and Berger, 1999; Narlikar et al., 2002).

It is thought that a specific chain of events involving transcription factor binding, histone modification, chromatin remodelling, polymerase recruitment and promoter firing occurs in the activation of every gene. However, very few studies have focused

on the order of events that occur in transcriptional activation. One exception to this is a recent report that provides an incisive analysis of the events occurring in *GADD45* induction by p53 (An et al., 2004).

p53 is a classical transcription activator with an N-terminal activation domain, a central DNA-binding domain and a C-terminal regulatory domain. It is stabilised following DNA damage and induces the transcription of a large number of direct target genes. *In vitro* assays using reconstituted chromatin and *in vivo* studies using chromatin immunoprecipitation demonstrated that – at the *GADD45* promoter – p53 recruits PRMT1, followed by p300, followed by CARM1 and that all three are needed for optimal activation (An et al., 2004). According to the histone code hypothesis, modification of histone tails provides a binding site for additional factors (Jenuwein and Allis, 2001) and this seems to be precisely what happens in the *GADD45* promoter. CARM1 methylates histone H4 on Arg3, facilitating the recruitment of p300, which acetylates H4 on multiple sites. This, in turn, facilitates the recruitment of CARM1, which acetylates histone H3 on Arg17. Recruitment of additional coactivators and resulting histone modifications were detected in this study but their position in the cascade of events leading to gene activation was not addressed (An et al., 2004).

These events probably occur as a prelude to chromatin remodelling at p53-dependent promoters. As a general principle, chromatin remodelling complexes are recruited either simultaneously with histone modifying enzymes or by direct interaction with specific histone modifications. The bromodomain of Brg1, for example, binds acetylated Lys8 on histone H4 and recruits the SWI/SNF complex (Hassan et al., 2002). p53 has also been shown to interact with the Mediator complex (Gu et al., 1999) and – though this study did not address it – this is probably important for recruiting the basal transcription apparatus after chromatin has been remodelled. Generally speaking, after chromatin has been opened, the basal transcription apparatus is recruited by a combination of direct interaction with coregulators and interaction with specific histone modifications. The double bromodomain of $TAF_{II}250$, for example, interacts with acetylated lysines and facilitates TFIID recruitment (Jacobson et al., 2000).

Once polymerase has been recruited and the promoter has fired, a final mechanistic challenge faced by the transcription machinery is how to elongate through less open chromatin. This is made possible by the FACT and Elongator complexes, which associate with RNAPII as it moves along the DNA and propagate the remodelling of chromatin throughout the ORF (Kim et al., 2002; Orphanides et al., 1998).

Gene repression and long-term silencing probably involve similarly ordered cascades of events. This is likely to involve establishment of a set of histone marks correlating with repression, followed by recruitment of chromatin assembly factors and DNA methyltransferases. One example of this type of mechanism is binding of the chromodomain of HP1 to dimethylated Lys9 of histone H3, leading to heterochromatin formation and long-term repression (Jacobs and Khorasanizadeh, 2002).

1.1.4 Signal Integration

Signalling pathways and the gene responses they initiate do not stand in isolation. Transcription factors activated by different pathways can interact on gene promoters and have combinatorial effects on transcription. Signalling pathways can also share common upstream components and activation of one pathway can promote or suppress the response to other signals. These points of contact between different pathways are collectively known as "crosstalk" and function to integrate the responses of the cell to different combinations of external cues (Dumont et al., 2001).

Signals may be integrated in a synergistic, cooperative or antagonistic fashion. For example, different transcription factors cooperate synergistically in viral induction of the *IFNβ* promoter (see above, Fig. 1.1) (Chen, 1999; Courey, 2001). Synergistic integration is also observed between ERK1 and GSK3 in the regulation of the heat-shock transcription factor HSF-1. ERK is activated by the canonical Ras-Raf-MEK-ERK cascade whereas GSK3 is activated by PKA downstream of glucagon and adrenaline. Sequential phosphorylation of HSF-1 by ERK followed by GSK3 is required to repress the ability of HSF-1 to induce target genes such as *Hsp70B*. It is

thought that these mechanisms function to ensure that the heat-shock response is suppressed during tissue growth (Chu et al., 1996).

An example of additive cooperation between signals is seen in induction of the gene encoding the α2 subunit of type VII collagen. TGF-β and TNF-α each induce this gene and co-stimulation with both ligands gives an elevated effect (Kon et al., 1999). An example of antagonistic signal integration is the repression of Ras-AP1 function by IFN-γ-Stat1 signalling. This is thought to result from Stat1 cooperating more effectively than AP-1 factors for CBP/p300 (Horvai et al., 1997). Another example is the crosstalk between IFN-γ and TGF-β in leukocytes. Stat1 is activated downstream of IFN-γ and transcriptionally induces the expression of Smad7, a strong negative regulator of TGF-β signalling. TGF-β inhibits leukocyte function and this mechanism of signal integration may allow IFN-γ to overcome the inhibitory effect of TGF-β and stimulate immune and inflammatory responses (Ulloa et al., 1999).

1.2 TGF-β and its biological functions

The transforming growth factor β (TGF-β) superfamily consists of a large group of secreted signalling molecules that are conserved throughout the animal kingdom and which regulate virtually every aspect of embryonic development and adult tissue homeostasis (Shi and Massague, 2003). This group includes members of the TGF-β, Activin/Nodal and BMP families. Signalling by the TGF-β superfamily involves the familiar system of a secreted protein binding to a cell-surface receptor and inducing or repressing the transcription of downstream target genes. This occurs via an apparently simple pathway in which members of the Smad family play a central role. However, intricate regulation at the molecular level and extensive crosstalk with other cellular factors allows cells to respond differently depending on their type and physiological status (ten Dijke and Hill, 2004).

1.2.1 Overview: TGF-β function and signalling mechanisms

Mammalian TGF-β1, TGF-β2 and TGF-β3 are the prototypical members of the TGF-β superfamily and one or more of these ligands is produced in almost every tissue at some stage in development (Pelton et al., 1991). These factors share a very high degree of structural and functional similarity (they probably differ mainly in their expression patterns) and, in this thesis; the term "TGF-β" will be used generically for all three. As in embryos, TGF-β is also ubiquitous molecule in adult tissues. Its function is highly context-specific and TGF-β is amongst the most pleiotropic of mammalian signalling factors (Siegel and Massague, 2003). In epithelia, for example, it inhibits cell cycle progression in G_1 (Cui et al., 1996; Laiho et al., 1990). In endothelial cells, however, TGF-β functions to stimulate proliferation, partly by induction of *Id1* (Goumans et al., 2003). Similarly, TGF-β inhibits proliferation at almost all stages of B and T cell development but at the same time is required for isotype switching in the B cell lineage (Gorelik and Flavell, 2002; Lebman and Edmiston, 1999).

In order to facilitate further explanation of TGF-β function, it is necessary to first give a brief overview of the signalling pathway downstream of this ligand. TGF-β signalling begins with release and activation of dimeric TGF-β. This then mediates the assembly of heterotetrameric receptors consisting of two type I (TβRI) and two type II (TβRII) TGF-β receptors on the surface of recipient cells. The constitutively active intracellular domain of TβRII then phosphorylates and activates TβRI. Active TβRI is thus enabled to bind and phosphorylate Smad2 and Smad3. These receptor-regulated Smads (R-Smads) then dissociate from the receptor and form heterodimeric and heterotrimeric complexes with one molecule the common mediator Smad (Co-Smad), Smad4. In both the presence and absence of TGF-β all these Smads are continually shuttling between cytoplasm and nucleus but, upon formation of active Smad complexes, they accumulate in the nucleus. In the nucleus, active Smad complexes associate with cell-type specific transcription factors on target gene promoters and induce or repress the expression of these genes (Fig. 1.8A). After several hours of signalling, various negative feedback mechanisms inactivate the

TGF-β receptors and the pathway is shut down (Fig 1.2) (Shi and Massague, 2003; ten Dijke and Hill, 2004).

Though they bind different type I and type II receptors, other members of the TGF-β superfamily signal via very similar downstream pathways to TGF-β itself. Smad2, Smad3 and Smad4 are the canonical signalling mediators of the Activin/Nodal family and function in the same way as in TGF-β signalling. BMP family members canonically use Smad1, Smad5 and Smad8 as R-Smads but, like TGF-β, also rely on Smad4 as their Co-Smad (Fig. 1.3) (Chen et al., 2004).

1.2.2 The TGF-β superfamily in embryonic development

Embryonic development is the context where both the biological functions of the TGF-β superfamily and the signalling mechanisms involved have been investigated most thoroughly. Generally speaking, members of the superfamily function to establish morphogen gradients that help set up the overall body plan of the embryo. They also function in establishment of the three germ layers and in the genesis of specific organs (Chang et al., 2002; Hill, 2001). The *Drosophila melanogaster* BMP homolog Decapentaplegic (Dpp), for example, functions in establishing the dorsoventral axis of the embryo. Dpp forms a gradient with activity levels highest on the dorsal side of the embryo and lowest on the ventral. High levels of Dpp specify dorsal fates while low levels allow ventral fates to be determined, thus establishing the dorsoventral axis of the fly (Morisato and Anderson, 1995).

Many members of the TGF-β superfamily are also active in early embryos of *Xenopus laevis*, including various BMPs, Activin, *Xenopus* nodal-related proteins (Xnrs) and the Activin/Nodal family members Derriere and Vg1. As in *Drosophila*, they function largely by forming morphogenetic gradients that supply cells with positional information required for appropriate differentiation and morphogenetic movements. BMP family members, for example, are required to specify the dorsoventral axis in *Xenopus*, just as in *Drospohila*. However, unlike in *Drosophila*, signal levels are highest on the ventral side and lowest on the dorsal side due to an inversion of axis

specification mechanims that appears to exist between vertebrates and invertebrates. Members of the Activin/Nodal family are also essential for early embryonic development in *Xenopus*. They are needed to specify mesoderm and endoderm and for proper gastrulation movements (Hill, 2001).

Nodal signalling is also involved in gastrulation movements and germ-layer patterning in zebrafish and mouse. *Nodal*$^{-/-}$ mice fail to specify mesoderm and their development arrests at gastrulation. Mice carrying a null allele and a hypomorphic allele of *Nodal* are able to gastrulate but show defects in anteriorposterior axis specification, anterior patterning and left-right asymmetry. Expression of the Nodal ligand trap Lefty on the left side of the embryo represses Nodal activity on that side and allows the left-right axis to be established. Smad-interacting transcription factors required for Nodal function in mouse and *Xenopus* embryos are also expressed in cell lines derived from early human embryos, suggesting that Nodal is also important in human development (Whitman, 2001).

As described above, TGF-β itself is also widely expressed in mouse embryos, albeit at later stages than BMP and Activin/Nodal family members (Pelton et al., 1991), and knockout studies reveal that TGF-β1, TGF-β2 and TGF-β3 all play essential roles in specific processes of organogenesis. Half of *TGF-β1*$^{-/-}$ mice die *in utero* from deficient yolk sac vascularisation and the remainder die within one month from severe inflammation (Dickson et al., 1995; Kulkarni et al., 1993; Martin et al., 1995). *TGF-β2*$^{-/-}$ mice show skeletal, craniofacial, cardiac and other defects and die shortly after birth (Sanford et al., 1997). *TGF-β3*$^{-/-}$ mice have a cleft palate and delayed pulmonary development and also die shortly after birth (Kaartinen et al., 1995; Proetzel et al., 1995). These non-overlapping phenotypes reveal that the three members of the mammalian TGF-β family have distinct biological functions. However, since these ligands are biochemically very similar, this is probably a result of their different expression patterns (Pelton et al., 1991), rather than major intrinsic differences between the molecules themselves.

1.2.3 TGF-β and cell cycle regulation

In embryonic tissues, the biological functions of TGF-β require it to be able to induce a wide range of cellular responses in different contexts. These include proliferation, migration, apoptosis and various programmes of differentiation. Many of these functions are maintained in adult tissues but the best-characterised and possibly most general function of TGF-β in this setting is to inhibit cell cycle progression in G_1. TGF-β has this cytostatic effect in epithelial cells, lymphocytes and other cell types.

Target gene responses involved in inhibiting G_1 progression include direct Smad-mediated induction of $p21^{Cip1/Waf1}$ and $p15^{INK4B}$ (Hannon and Beach, 1994; Reynisdottir et al., 1995). Induction of $p21^{Cip1/Waf1}$ occurs in cooperation with FoxO, p53 and possibly other factors (Cordenonsi et al., 2003; Seoane et al., 2004) whereas induction of $p15^{INK4B}$ has been reported to occur in cooperation with Sp1 (Feng et al., 2000). Active Smad complexes also directly repress *c-Myc*, *Cdc25*, *Id1*, *Id2* and *Id3* (Iavarone and Massagué, 1997; Kang et al., 2003; Seoane et al., 2001). Repression of *c-Myc* occurs in cooperation with E2F4/5 and the corepressor p107 (Chen et al., 2002b). c-Myc is a bHLH-leucine zipper transcription factor that binds the $p15^{INK4B}$ promoter in association with the zinc finger protein Miz-1. This complex represses $p15^{INK4B}$ transcription and, as a result, repression of *c-Myc* by TGF-β synergises with the direct effect of Smad complexes on $p15^{INK4B}$ expression.

p21 is a CDK inhibitor (CKI) that can bind Cdk4/6-Cyclin D, Cdk2-Cyclin E and Cdk2-Cyclin A complexes to inhibit their catalytic activity (Lee and Yang, 2001). It can also act as a scaffold for the formation of Cdk4/6-Cyclin D complexes. How the switch is regulated is poorly understood, but this function of p21 does not appear to be relevant in the context of TGF-β signalling (Siegel and Massague, 2003). p15 is also a CKI but in this case with specificity only for Cdk4/6-Cyclin D complexes (Lee and Yang, 2001). Interestingly, binding of p15 to these complexes displaces p21, releasing it to block Cdk2-Cyclin E complexes and thus enhancing the cytostatic effect of TGF-β (Reynisdottir et al., 1995).

In response to mitogenic signals, c-Myc helps induce expression of D-type cyclins, Cdk4, Cdk6 and E2F. Similarly, Cdc25A is a phosphatase required to remove an inhibitory phosphate from near the active site of Cdk4 and Cdk6 so that they can be activated by D-type cyclins (Hoffmann et al., 1994). Id1, Id2 and Id3 are HLH proteins that promote cell cycle progression by releasing E2F from the effects of Rb (Lasorella et al., 2000) as well as by inhibiting differentaition. Along with induction of CKIs, downregulation of all these factors by TGF-β contributes to its negative effect on G_1 progression (Siegel and Massague, 2003).

TGF-β may not inhibit the cell cycle at all in endothelial cells. Instead it seems to actually stimulate proliferation. As noted already, this is probably dependent on induction of *Id1* by Smad1/5 signalling (Goumans et al., 2003). However, other target genes are also likely to be involved. Future studies may reveal more settings in which TGF-β can stimulate cell proliferation.

1.2.4 TGF-β and apoptosis

TGF-β has been reported to induce apoptosis in mammalian cells (Siegel and Massague, 2003). However, the extent to which this occurs *in vivo* remains unclear; it seems to depend greatly on cell type and in some cases the effect is clearly indirect. Possible mechanisms include TGF-β-dependent downregulation of Bcl-XL or upregulation of Caspase 3 and Caspase 8 (Chen and Chang, 1997; Saltzman et al., 1998). In addition, certain well-characterised Smad target genes are known to play a role in apoptosis. *TIEG1*, for example encodes a transcription factor that has been shown to induce apoptosis in several cell types, probably by repressing *Bcl-2* and *Bcl-XL* (Tachibana et al., 1997). Another TGF-β target gene, *SHIP* encodes an inositol-5-phosphatase that inhibits activation of the survival kinase PKB/Akt by PDK1 (Valderrama-Carvajal et al., 2002). Finally, the adaptor protein Daxx – which is involved in activation of Caspase-8 and Caspase-10 by Fas receptors – has been reported to interact with TβRII and seems to be necessary for TGF-β-induced apoptosis (Perlman et al., 2001).

Whether all of these mechanisms operate in different cell types or are part of a more general network remains unclear. Likewise, little is known about whether changes in the ability of TGF-β to induce apoptosis play an important role in cancer (Siegel and Massague, 2003). However it has been shown that sustained activation of Raf in MDCK cells induces TGF-β expression while inhibiting its apoptotic response but not affecting its invasive responses. A similar process could well be important in Raf- or Ras- transformed carcinoma cells (Lehmann et al., 2000).

1.2.5 TGF-β, cellular invasion/migration and EMT

In certain cell lines, particularly Ras-transformed epithelial cells, TGF-β stimulates increased migration and enables the cells to penetrate collagen matrices (Oft et al., 1998) (Hill, C.S. and Daly, A.D., unpublished data). Gene responses involved in these functions of TGF-β include induction of the matrix metalloproteases MMP-2 and MMP-9, induction of PAI-1 and repression of the MMP inhibitor TIMP (Duivenvoorden et al., 1999; Kordula et al., 1992; Laiho et al., 1986; Sehgal and Thompson, 1999). Similar effects probably also occur *in vivo*, as suggested by the ability of anti-TGF-β1 antibodies to inhibit cyclosporin-induced metastasis of A549 cells xenografted into nude mice (Hojo et al., 1999).

Canonical EMT (epithelial to mesenchymal transition) involves a switch from an epithelial to a mesenchymal morphology. This is accompanied by an increase in the motility and invasiveness of the cell. Molecular events underlying this process include downregulation of the epithelial markers E-cadherin, ZO-1, vinculin and cytokeratin along with upregulation of N-cadherin and vimentin (Thiery, 2003). In cultured cells where Ras is activated, TGF-β has a well-documented ability to induce these changes (Oft et al., 1998) (Hill, C.S. and Daly, A.D., unpublished data) and there is evidence that TGF-β can induce EMT *in vivo* (Cui et al., 1996; Portella et al., 1998).

One way in which TGF-β induces EMT is by induction of Snail, a zinc-finger transcription factor that, along with the related protein Slug, acts as a master regulator

of EMT in mouse gastrulation (Thiery, 2003). Snail directly represses the genes encoding E-Cadherin, Cytokeratin 18 and other epithelial markers (Cano et al., 2000; Nieto, 2002). It also indirectly induces Smad-interacting protein 1 (SIP1)/ZEB2, another zinc-finger factor that represses E-cadherin expression (Guaita et al., 2002). Whether continued TGF-β signalling and interaction with active Smad complexes is important for this function of SIP1 has not been investigated. Repression of *Id2* by TGF-β also appears to be important, since Id2 is a repressor of the gene encoding α-SMA (smooth muscle actin), a marker of EMT in certain systems. Knock-down of Id2 has been shown to be sufficient to allow BMP7 to induce EMT in NMuMG and NMe cells (Kondo et al., 2004).

1.2.6 TGF-β and angiogenesis

The TGF-β pathway is involved in both normal and pathological angiogenesis. Mouse knockouts and naturally occurring human mutations show that TGF-β, TβRI, TβRII and various other components of TGF-β receptor complexes (ALK1 and endoglin – see section 1.3 below) are all required for normal vasculogenesis (Chang et al., 2002). In accordance with this, TGF-β has been shown to induce capillary formation by endothelial cells grown on a collagen matrix (Madri et al., 1988). TGF-β directly stimulates proliferation and migration of endothelial cells (Madri et al., 1988) and also induces the expression of VEGF, which is well known as a powerful inducer of angiogenesis (Pertovaara et al., 1994). TGF-β also attracts monocytes, which produce various pro-angiogenic cytokines (Ashcroft, 1999).

1.2.7 TGF-β and the immune system

In the immune system, TGF-β inhibits proliferation and terminal differentiation of B cells, T cells, NK cells and macrophages (Gorelik and Flavell, 2002; Lebman and Edmiston, 1999). TGF-β directly inhibits proliferation of immune cells and also downregulates both IL-2 and the IL-2 receptor, critical components in the activation of cytotoxic T lymphocytes (CTLs) (Kehrl et al., 1986a; Kehrl et al., 1986b). TGF-β

also downregulates expression of MHC class II molecules and may make malignant and virus-infected cells intrinsically less immunogenic (Czarniecki et al., 1988).

1.2.8 TGF-β and cutaneous wound healing

Immunomodulation is one important physiological role of TGF-β, another is regulation of cutaneous wound healing (Shah et al., 1994). In this process, platelets release large amounts of TGF-β to stimulate migration of keratinocytes into the wound as well as matrix production by fibroblasts and movement of monocytes into the affected area. The initial burst of TGF-β released by platelets also establishes autocrine loops in other cell types, causing them to produce lower levels of TGF-β for longer time periods (Massague, 1999). Interestingly, it has been shown that Smad3-null mice have accelerated healing of incisional wounds, suggesting that the role of TGF-β in wound healing is complex (Ashcroft et al., 1999). It is known that loss of Smad3 attenuates the cytostatic effect of TGF-β on keratinocytes and it should also inhibit the matrix-inducing effect on fibroblasts. It has been proposed that, in combination, these effects would result in accelerated formation of scar tissue that contains elevated levels of keratinocytes and reduced levels of matrix. Such scar tissue would probably be less robust than normal scar tissue. The function of TGF-β in wound healing may thus be to ensure the formation of robust scar tissue by restraining keratinocyte proliferation while inducing matrix formation and migration of cells into the wound (Massague, 1999).

1.2.9 TGF-β and disease

In accordance with the diversity of functions it displays in normal physiological function, TGF-β has been linked to a wide range of human diseases. TGF-β overexpression has been shown to contribute to a range of fibrotic conditions. These include renal fibrosis induced by a various insults, hepatic fibrosis associated with hepatitis C infection and postoperative scarring of the cornea. These findings fit with the ability of TGF-β to induce extracellular matrix production. Another disease

associated with TGF-β is hereditary haemorrhagic telangectasia, a congenital disorder involving defective blood vessel patterning and caused by genetic defects in endoglin and ALK1 (see section 1.2.6 above and section 1.3 below) (Blobe et al., 2000).

TGF-β is also associated with atherosclerosis, where it plays a protective role by inhibiting proliferation of smooth muscle cells and migration of endothelial cells (Blobe et al., 2000). However, it has been shown that high levels of TGF-β can contribute to hypertension in mice, suggesting a complex relationship between TGF-β and vascular disease (Zacchigna et al., 2006). Finally, viral infection can modulate TGF-β function. The pathologies associated with HIV and HBV involve amplification of TGF-β signalling while those associated with HTLV and KSHV (both of which are oncogenic) involve antagonism of TGF-β function (Hall and Fujii, 2005; Pan et al., 2004; Reinhold et al., 1999; Tomita et al., 2004).

The pathology in which the role of TGF-β has been investigated most extensively is cancer. In this disease, TGF-β seems to play a complex role – acting as both a tumour suppressor and a tumour promoter in different contexts. Despite having initially been isolated from sarcomas (Anzano et al., 1983), TGF-β was for many years described mainly as a tumour suppressor. The signalling pathway downstream of TGF-β has been shown to undergo loss of function in many tumour types and this is known to contribute to tumour growth (Levy and Hill, 2006). However, TGF-β is overexpressed by a majority of malignant carcinomas – including those originating in the breast, colon, stomach, pancreas and prostate – and can have strong tumour-promoting effects (Levy and Hill, 2006).

How can these two observations be reconciled? In the past decade a two-step model has emerged in which TGF-β generally acts as a tumour suppressor in early stages of oncogenesis and as a tumour promoter later on (Cui et al., 1996). At early stages in tumourigenesis, loss of function in the TGF-β pathway leads to loss of the cytostatic and apoptotic responses to this cytokine and enhanced tumour growth. However, cancers that do not undergo loss of function in the TGF-β pathway itself can lose sensitivity to its cytostatic effects by other means and may then overproduce this cytokine. Tumours overexpressing TGF-β are then at a selective advantage due to its

ability to inhibit local immune function and stimulate migration, invasion and angiogenesis (Derynck et al., 2001).

Genetic changes affecting several components of the TGF-β pathway contribute to early stages of oncogenesis. In the case of Smad4, a combination of loss of heterozygosity and intragenic mutation in the other allele occurs frequently in pancreatic cancer, sporadic colon cancer and breast cancer. Similarly, wild-type TβRII is frequently lost in sporadic colon cancer and breast cancer (Levy and Hill, 2006). Interestingly, a subset of TGF-β gene responses have been shown to be Smad4-independent, raising the possibility that these are retained in Smad4-null tumours where they contribute to tumour progression (Levy and Hill, 2005).

However, "epigenetic" changes affecting the cytostatic response to TGF-β are probably more common that genetic changes affecting the pathway itself. Such "epigenetic" changes leave the core TGF-β–Smad pathway intact, allowing it to mediate various tumour-promoting effects, while eliminating effects of TGF-β on the cell cycle (Akhurst and Derynck, 2001). One event that is likely to contribute to this is loss of Rb, which occurs in most tumours. Since the best-characterised effects of TGF-β on the cell cycle all appear to operate via Rb (Siegel and Massague, 2003), it seems likely that loss of this tumour suppressor must affect the cytostatic response to TGF-β. Similarly, p53 has been shown cooperate with active Smad complexes on the $p21^{Cip1/Waf1}$ promoter and is required for induction of this gene by TGF-β (Cordenonsi et al., 2003). Since p53 is lost in over half of all cancers, this may provide another general mechanism whereby the cytostatic response to TGF-β is attenuated.

Deregulated expression of c-Myc is a further means by which tumour cells can overcome the effects of TGF-β on the cell cycle (Chen et al., 2001). It has been shown that a repressive complex containing Smad3 and Smad4 that normally forms on the proximal promoter of *c-Myc* following TFG-β treatment (Warner et al., 1999) does not form in certain breast cancer cell lines. This prevents TGF-β-dependent repression of *c-Myc*. In these cells, active Smad complexes formed normally on the promoters of other TGF-β target genes, suggesting that a cofactor essential for binding the *c-Myc* promoter is either missing or downregulated (Chen et al., 2001).

Other ways in which *c-Myc* could escape repression by TGF-β include translocation, gene fusion, gene amplification and promoter mutation.

In a similar fashion, the cytostatic response to TGF-β seems to be attenuated in some (but not all) Ras-transformed epithelial cells. This is may be due to ERK-mediated phosphorylation of sites in the linker regions of Smad2 and Smad3 and reduced nuclear accumulation of active Smad complexes (Kretzschmar et al., 1999). However, it is possible that downstream effects of Ras signalling could block the function of cofactors required for specific TGF-β gene responses. This is supported by the finding that Ras activation has no direct effect on Smad function in some cell lines (Lehmann et al., 2000).

Much remains unknown about the mechanisms by which tumour cells evade the cytostatic effects of TGF-β. It is an area that is under extensive investigation but progress has been hampered by limitations in our understanding of Smad-cofactor interactions. Greater knowledge of the transcription factors and coregulators that are required for specific TGF-β gene responses must be accumulated before it becomes possible to understand how cells can lose some responses to TGF-β but not others.

1.3 TGF-β–Smad signalling: structures and mechanisms

In the past decade, intense investigation has yielded a detailed picture of the molecular processes that transmit the TGF-β signal from the cell membrane to the nucleus. This ongoing research has relied on detailed characterisation of the individual components of the pathway as well as of their physical and functional interactions. In this section I will summarise the conclusions of this work by presenting what is known about particular sets of pathway components (ligands, receptors and Smad molecules) and immediately afterwards describing the steps of the pathway that they mediate. The overall dynamics of the pathway are summarised in Fig. 1.2.

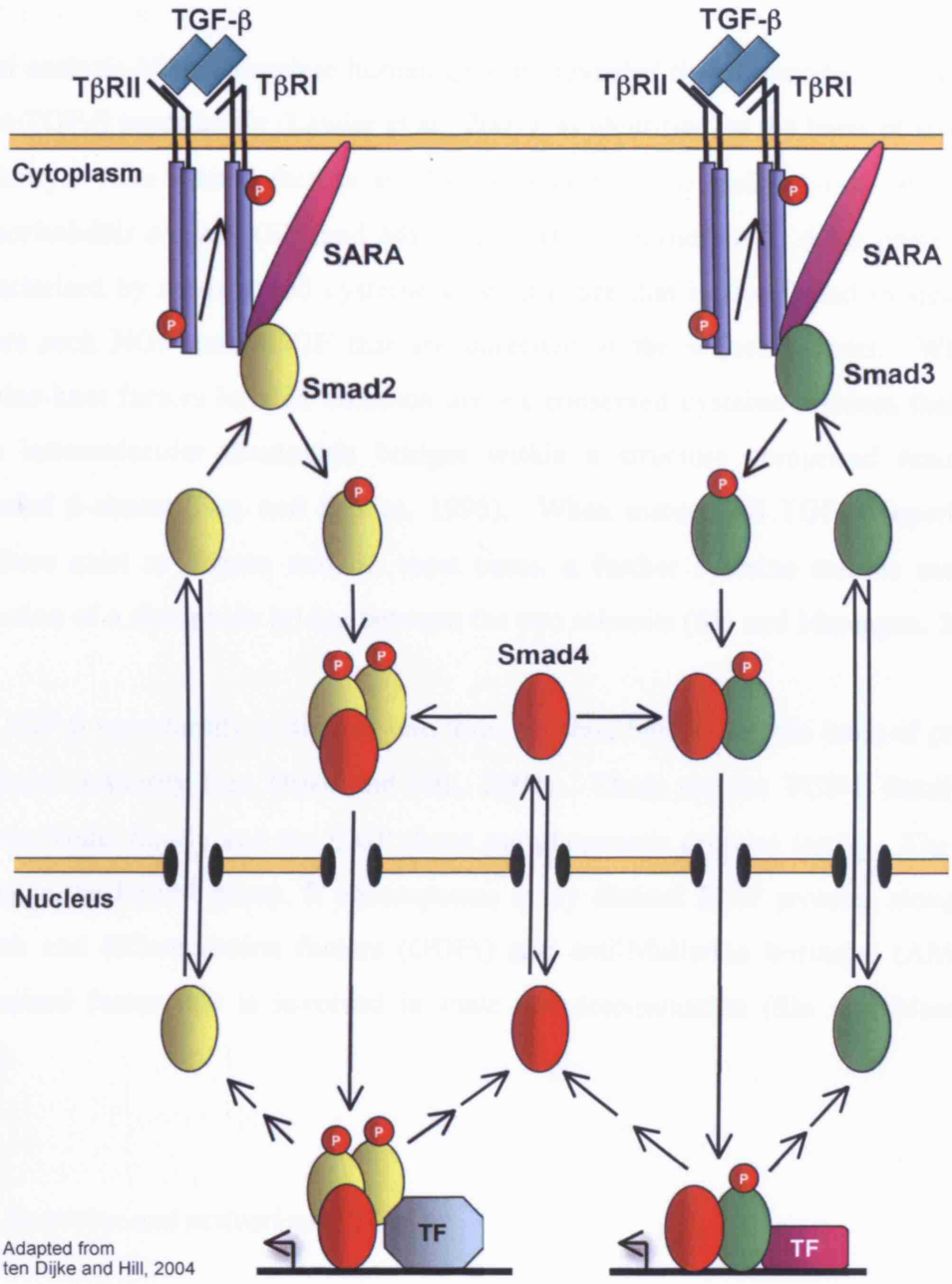


Fig. 1.2 The TGF-β pathway and Smad nucleocytoplasmic shuttling. Dimeric TGF-β brings together TβRI and TβRII homodimers, allowing TβRII to phosphorylate and activate TβRI. Smad2 and Smad3 are recruited by SARA and are phosphorylated by active TβRI. They then form heterodimers and heterotrimers with Smad4. These accumulate in the nucleus, associate with transcription factors (TF) and bind target gene promoters to control gene expression. In complexes whose stoichiometry has been determined, Smad2 forms heterotrimers with Smad4 whereas Smad3 forms heterodimers (Inman and Hill, 2002). Active Smad complexes are inactivated by unknown mechanisms and Smads are returned to the cytoplasm, where Smad2 and Smad3 can be phosphorylated again, allowing the formation of new Smad complexes. In the absence of TGF-β, Smad molecules are also continually shuttling between the cytoplasm and nucleus, but are more abundant in the cytoplasm.

1.3.1 The TGF-β superfamily

Initial analysis of the complete human genome revealed that it encoded 42 members of the TGF-β superfamily (Lander et al., 2001), as identified on the basis of sequence similarity. Nine related factors are found in *Drosophila melanogaster* and six in *Caenorhabditis elegans* (Shi and Massague, 2003). Structurally, these proteins are characterised by a conserved cysteine-knot structure that is also found in signalling factors such NGF and PDGF that are unrelated at the sequence level. What all cysteine-knot factors have in common are six conserved cysteine residues that form three intramolecular disulphide bridges within a structure comprised mainly of extended β-sheets (Sun and Davies, 1995). When mature, all TGF-β superfamily members exist as dimers and, in most cases, a further cysteine residue mediates formation of a disulphide bridge between the two subunits (Shi and Massague, 2003).

The TGF-β superfamily is divided into three separate families on the basis of primary sequence similarity (ten Dijke and Hill, 2004). These are the TGF-β family, the Activin/Nodal family and the BMP (bone morphogenetic protein) family. The BMP family is the largest group. It encompasses many distinct BMP proteins along with growth and differentiation factors (GDFs) and anti-Mullerian hormone (AMH), a specialised factor that is involved in male sex determination (Shi and Massague, 2003).

1.3.2 Secretion and activation of TGF-β

TGF-β is secreted as an inactive complex in which the two subunits of the active form are noncovalently associated with the prodomians that have been cleaved from each subunit in the secretory pathway. This complex also contains a large protein known as latent TGF-β binding protein (LTBP) that enables proper assembly and secretion of TGF-β (Taipale et al., 1998). Inactive TGF-β complexes – which are sequestered by the extracellular matrix (ECM) – are activated by the proteases plasmin, MMP-2 and MMP-9 as well as by thrombospondin and αvβ6 integrin (Crawford et al., 1998; Lyons et al., 1990; Munger et al., 1999; Stamenkovic, 2000; Yu and Stamenkovic,

2000). Access of TGF-β1 to its cognate receptors is regulated by a range of accessory factors that include the coreceptors betaglycan and endoglin and the ligand traps decorin and α2-macroglobulin (Brown et al., 1999; Lebrin et al., 2004; Shi and Massague, 2003). Unusually, connective tissue growth factor (CTGF) has also been shown to promote TGF-β binding to its receptors (Abreu et al., 2002).

1.3.3 The TGF-β receptor family

TGF-β receptors are all transmembrane protein serine/threonine kinases (Manning et al., 2002). All members of this kinase family function as receptors for the TGF-β superfamily; seven as type I receptors and five as type II receptors (Shi and Massague, 2003). The type II receptors are TβRII, ActRIIA, ActRIIB, BMPRII and AMHRII. TβRII binds the TGF-β family; ActRIIA and ActRIIB bind the Activin/Nodal and BMP/GDF families; BMPRII binds the BMP/GDF family and AMHRII binds AMH (Fig. 1.3). The type I receptors comprise the Activin receptor-like kinase (ALK) family. ALK5/TβRI is the canonical type I receptor for TGF-β but it can also signal via ALK1 (Figs. 1.3 and 1.4); ALK4 and ALK7 are the type I receptors for the Activin/Nodal family, ALK3 and ALK6 are the type I receptors for the BMP/GDF family and ALK2 is the type I receptor for AMH (Fig. 1.3). It is thought that binding to different combinations of receptors provides a means to diversify downstream responses (ten Dijke and Hill, 2004).

Structurally, all members of the TGF-β receptor family share a similar intracellular protein serine/threonine kinase domain that contains a canonical protein kinase fold. Type I receptors also contain the juxtamembrane GS motif that adopts a helix-loop-helix conformation and which is the site of phosphorylation by the type II receptor (Fig. 1.5A) (Huse et al., 2001). Similarly, the ectodomains of several family members have been shown to share a three-finger "toxin fold" in which each finger consists of a pair of antiparalell β strands (Boesen et al., 2002; Greenwald et al., 1999; Kirsch et al., 2000).

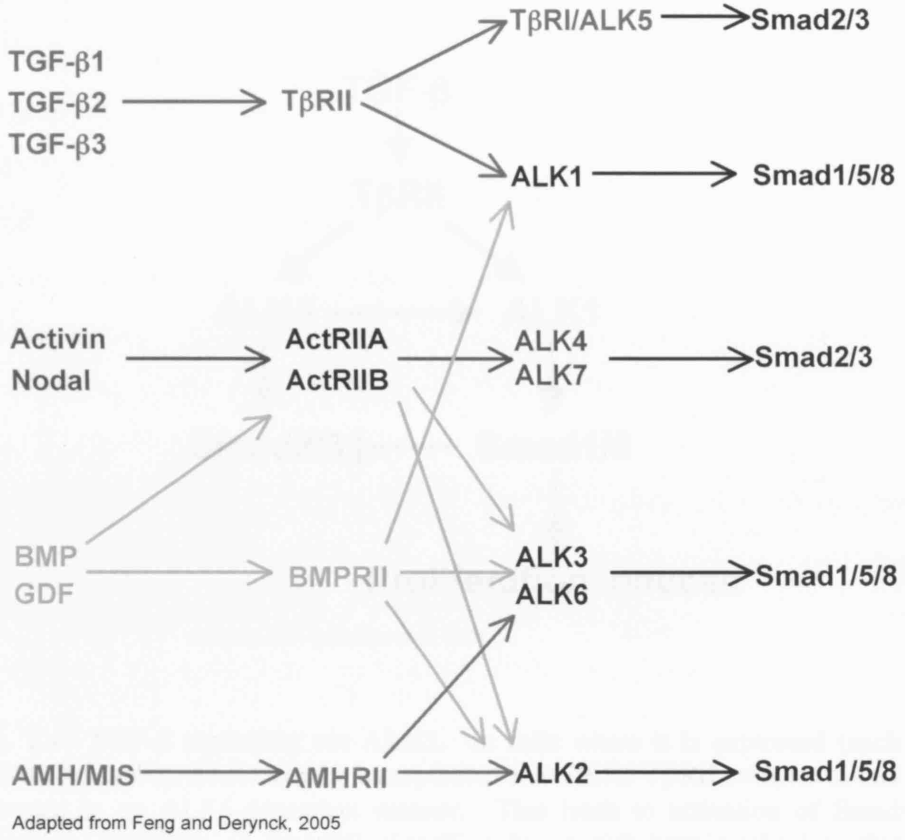
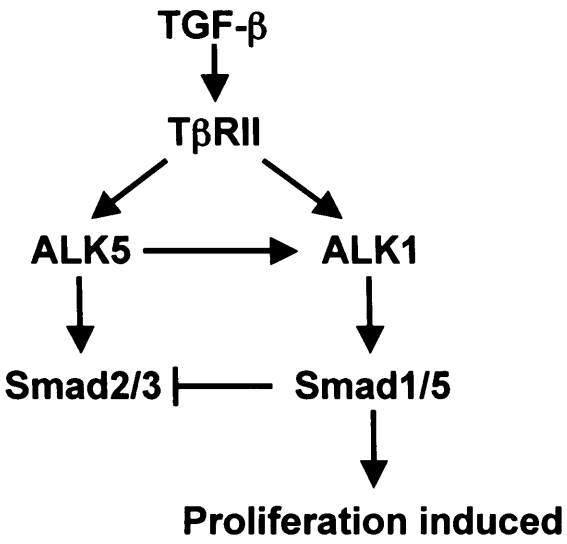


Fig. 1.3 Receptor and Smad activation by the TGF-β superfamily. TGF-β superfamily members are colour-coded in red, blue, orange and green. Receptors and signalling events (represented by arrows) specific to each ligand family are shown in a matching colour. Pathway components and signalling events that are shared by more than one ligand family are shown in black. Note that each ALK family member seems to activate the same Smads regardless of how the ALK itself was activated.



Adapted from Goumans *et al*, 2001.

Fig. 1.4 TGF-β signalling via ALK1. In cells where it is expressed (such as endothelial cells),, ALK1 can be incorporated into ALK5-TβRII complexes and be activated in an ALK5-dependent manner. This leads to activation of Smad1/5 signalling, inhibition of Smad2/3 signalling by an unknown mechanism that is dependent on Smad1 and stimulation of proliferation. This is consistent with stimulation of proliferation by BMP-activated Smad1 and Smad5.

1.3.4 Mechanisms of receptor activation

As mentioned above, TGF-β signals through a receptor complex that contains two distinct subunits, the type I TGF-β receptor (TβRI) and the type II TGF-β receptor (TβRII) (Yamashita et al., 1994). Independently of signal, both of these receptors exist as homodimers that are formed in the endoplasmic reticulum and persist in the plasma membrane (Gilboa et al., 1998). TGF-β has a higher affinity for TβRII dimers and binds these first, leading to a large conformational change in the ligand that enhances its affinity for the type I receptor (Hart et al., 2002; Rodriguez et al., 1995). This allows incorporation of TβRI dimers into heteromeric ligand-receptor complexes. (BMP on the other hand binds its type I receptors first, creating a complex that then recruits type II receptors (Shi and Massague, 2003).) Both type I and type II receptors have serine/threonine kinase activity in their intracellular domains (ten Dijke and Hill, 2004). TβRII, which autophosphorylates independently of TGF-β (Luo and Lodish, 1997), is thought to be constitutively active and, upon ligand stimulation, phosphorylates TβRI on one threonine and three serines in its juxtamembrane GS motif (Luo and Lodish, 1996; Shi and Massague, 2003; Weis-Garcia and Massagué, 1996).

Phosphorylation by TβRII causes a conformational change that moves the GS motif away from the adjacent L45 loop of the TβRI intracellular domain. This allows the GS motif to dock with a basic pocket located near the L3 loop on the surface of the MH2 domain of Smad2 and Smad3 and allows the L45 loop to interact directly with the L3 loop. Once the Smad-receptor interaction is established, the kinase domain of TβRI phosphorylates and activates Smad2 and Smad3 (Fig. 1.5A) (Huse et al., 2001; Shi and Massague, 2003). It is mainly through the L45-L3 interaction that type I receptors in general recognise the correct R-Smads to phosphorylate (Chen et al., 1998; Lo et al., 1998).

FKBP12 is a cytoplasmic protein that regulates receptor activation by binding the GS motif of TβRI and inhibiting receptor activation. In its unphosphorylated condition, the intracellular domain of TβRI exists in an equilubrium between an "open" or catalytically competent state – where ATP can gain access to the active site – and a

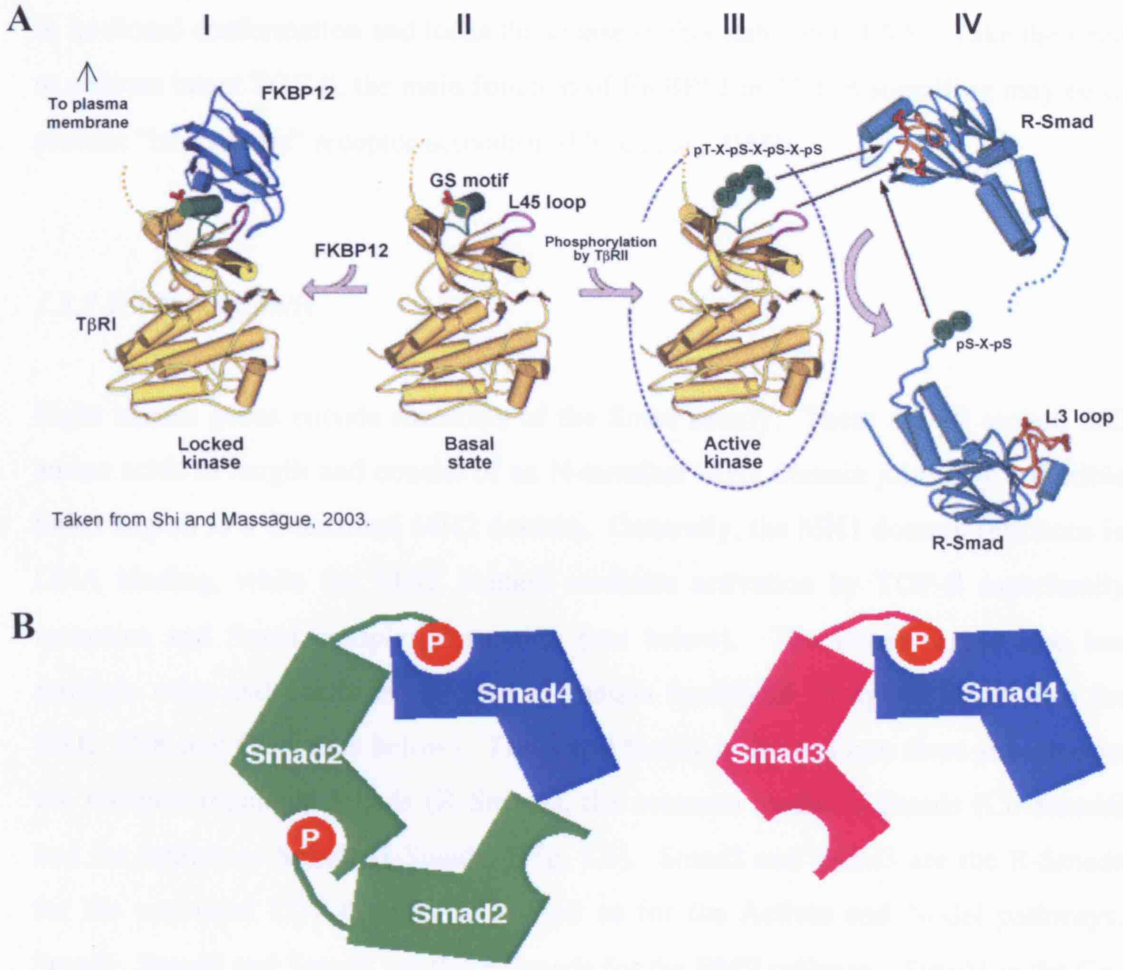


Fig. 1.5 Mechanisms of TβRI activation and Smad complex formation. *A) Activation of Tβ RI and R-Smads.* In the absence of signal, TβRI exists in an equilibrium between catalytically competent (II) and inactive (I) states. FKBP12 binds the GS motif of the inactive conformation and "locks" the kinase domain in this state (I). Upon addition of TGF-β, TβRI phosphorylates the GS motif (it is not known whether there is an active mechanism for displacing FKBP12) and stimulates formation of the active conformation. Phosphorylation of the GS motif also causes it to move away from the L45 loop of the receptor (III). The GS motif then binds a basic pocket on the MH2 domain of R-Smads and the L45 loop binds the Smad L3 loop adjacent to this pocket. With the kinase domain "locked" in the active conformation and surfaces created for R-Smad binding, these Smads are recruited and phosphorylated on serine residues at their extreme C-termini. Phosphoserine residues on one R-Smad molecule then bind another R-Smad in the same basic pocket as the phosphorylated GS motif of the receptor. This competitively disrupts the Smad-receptor interaction and mediates the formation of active Smad complexes (IV). It is not clear when Smad4 is incorporated or how R-Smad–Smad4 dimers form. *B) Schematic of dimers and trimers formed by Smad MH2 domains.* MH1 domains are omitted for clarity. The C-terminal phosphopeptide of R-Smads binds the basic pocket on an adjacent Smad molecule. Complexes typically contain one Smad4 molecule but complexes lacking Smad4 probably also form. Smad2 forms heterotrimers with Smad4 whereas Smad3 can also form heterodimers. Complexes containing both Smad2 and Smad3 also exist.

"closed" state where it cannot. FKBP12 associates with the GS motif while TβRI is in its closed conformation and locks the kinase in this state (Fig. 1.5A). Like the need to activate latent TGF-β, the main function of FKBP12 in TGF-β signalling may be to prevent "inadvertent" receptor activation (Huse et al., 1999).

1.3.5 The Smad family

Eight human genes encode members of the Smad family. These are all around 500 amino acids in length and consist of an N-terminal MH1 domain joined by a flexible linker region to a C-terminal MH2 domain. Generally, the MH1 domain functions in DNA binding, while the MH2 domain mediates activation by TGF-β superfamily receptors and Smad complex formation (see below). The linker region also has multiple roles and has been reported to contain functional phosphorlyation sites for ERK, JNK and CDK (see below). The Smad family is divided into three sub-groups: the receptor-regulated Smads (R-Smads), the common mediator Smads (Co-Smads) and the inhibitory Smads (I-Smads) (Fig. 1.6). Smad2 and Smad3 are the R-Smads for the canonical TGF-β pathway as well as for the Activin and Nodal pathways. Smad1, Smad5 and Smad8 are the R-Smads for the BMP pathway. Smad4 is the Co-Smad for all Smad signalling pathways. Smad6 and Smad7 are I-Smads that bind type I receptors in a similar fashion to R-Smads but which cannot be phosphorylated and instead act as negative regulators of the pathway (see below) (Shi and Massague, 2003).

X-ray crystallography shows that the MH2 domain – which is highly conserved amongst all Smads – consists of a central sandwich of two twisted β sheets capped at one end by a three-helix bundle and at the other by a loop/helix region (Fig. 1.5A) (Wu et al., 2001). The MH1 domain of the R-Smads is also highly conserved and consists of a compact structure in which an N-terminal half consists of three α helices and the C-terminal half of an α helix, six β strands, various loops and a projecting β hairpin that mediates interaction with DNA (Fig. 1.7A, see below) (Shi et al., 1998). At the sequence level, the MH1 domain of the I-Smads is quite divergent from that of the other family members (Shi and Massague, 2003).

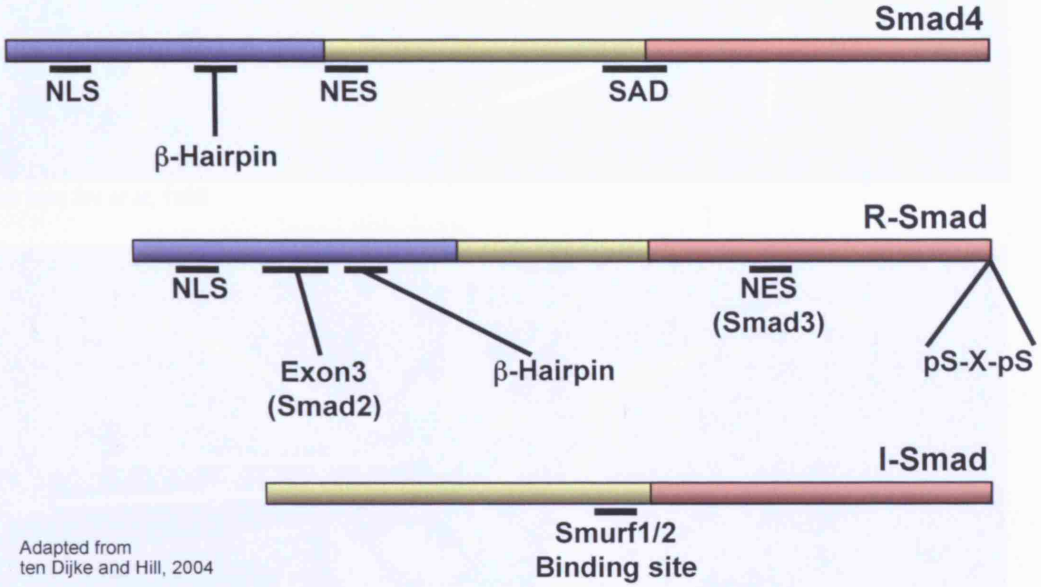
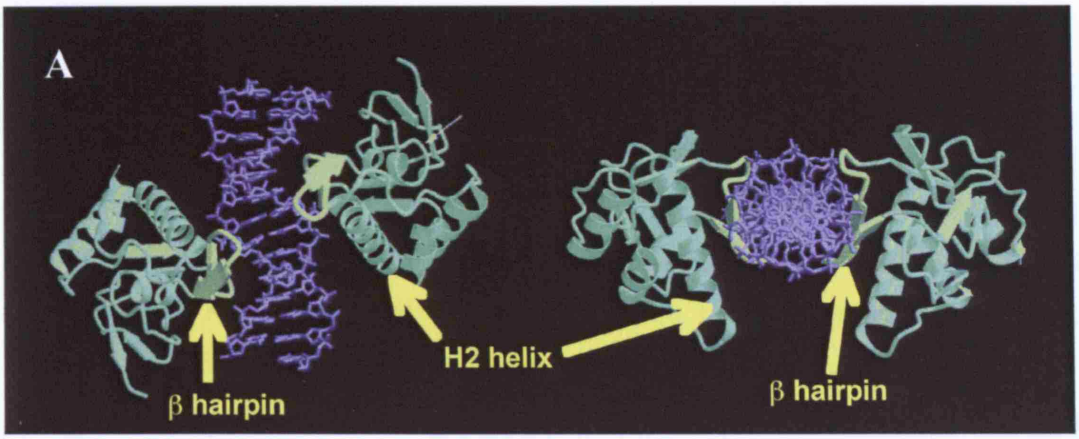
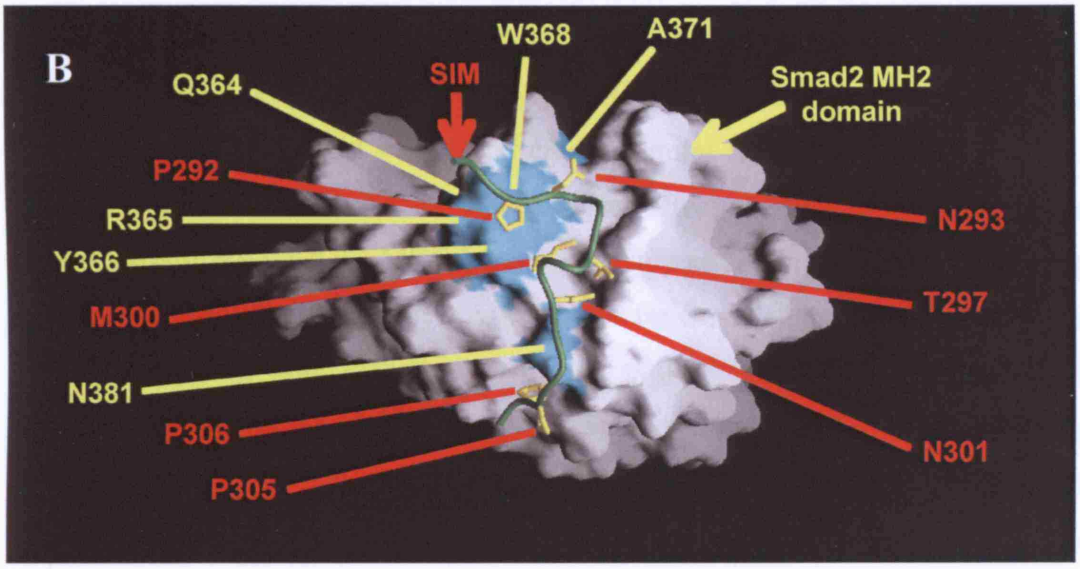


Fig. 1.6 Schematic alignment of Smad4, an R-Smad and an I-Smad. The MH1 and MH2 domains (blue and pink, respectively) and critical sequence features are shown. NLS: nuclear localisation signal, binds components of the importin α/β system. β-hairpin: Smad DNA-binding motif. NES: nuclear export signal, binds CRM1 in Smad4 and exportin 4 in Smad3. The NES of Smad2 has not been mapped. SAD: Smad activation domain, binds CBP/p300. Exon 3 (Smad2), alternatively spliced exon that is retained in the predominant isoform and which interferes with DNA binding. pS-X-pS: phosphorylated C-terminus of Smad2 (pS-M-pS) or Smad3 (pS-V-pS). Smurf1/2 binding site, region through which I-Smads interact with these E3 ubiquitin ligases to recruit them to active TGF-β receptors.



Adapted from Shi *et al*, 1998.



Adapted from Randall et al, 2002

Fig. 1.7 Smad structures. *A) Crystal structure of the Smad3 MH1 domain bound to DNA.* Two subunits are shown bound to distinct AGAC sites. The β-hairpin motifs are shown in yellow projecting into the major groove over these sites. Sequence-specific contacts are made with the first, second and fourth base-pairs in each AGAC site. Note that the H2 helix also lies adjacent to DNA but does not make direct contacts in this structure. *B) Structural model of the interaction between the Smad interaction motif (SIM) of* Xenopus *Mixer with the MH2 domain of human Smad2.* Critical residues in Smad2 are shown in cyan with yellow text and critical residues in the SIM are shown in yellow with red text.

1.3.6 Mechanisms of Smad activation

Prior to activation, Smad2 and Smad3 are recruited to the plasma membrane by the Smad anchor for receptor activation (SARA), a cytoplasmic protein containing a FYVE domain that binds phosphatidylinositol-3-phosphate in the membrane (Fig.1.2) (Tsukazaki et al., 1998). SARA also interacts with TβRI and it has been shown that activation of SARA-associated receptor complexes leads to their internalisation via Clathrin-coated pits to create signalling endosomes upon which Smad2 and Smad3 are activated (Di Guglielmo et al., 2003).

Phosphorylation of Smad2 and Smad3 by TβRI occurs on the second and third serines of an SSXS motif located at their extreme C-termini (Fig. 1.6) (Macías-Silva et al., 1996; ten Dijke and Hill, 2004). One proposed effect of this is to reduce the affinity of the MH2 domain for the MH1 domain, relieving an autoinhibition that holds the R-Smads in a conformation that is not permissive for Smad oligomerisation (Hata et al., 1997). Phosphorylation also creates an acidic projection that binds with high affinity to the L3 pocket on the surface of other R-Smad molecules and to a related (but less basic) pocket near the L3 loop of Smad4 (Fig. 1.5A). This interaction is necessary for the formation of Smad heterodimers and heterotrimers (Fig. 1.5B) (Chacko et al., 2001).

Phosphorylation of Smad2 and Smad3 also result in their dissociation from both SARA and the receptor. It is likely that this occurs by competitive mechanisms. Formation of Smad complexes in which the basic pocket near the L3 loop of Smad2 and Smad3 are occupied by the phosphorylated C-terminal serines of another R-Smad probably disrupts the Smad-receptor interaction and causes R-Smads to dissociate from the receptor (Fig. 1.5A) (ten Dijke and Hill, 2004; Wu et al., 2001). In addition, a region of the MH2 domains of Smad2 and Smad3 known as the β1' strand is involved in interactions with both SARA and other Smad molecules in a Smad dimer or trimer. This may result in competitive disruption of the R-Smad-SARA interaction and dissociation of active R-Smads from SARA (ten Dijke and Hill, 2004; Wu et al., 2000). It has been found that phosphorylated R-Smads undergo nuclear accumulation even in the absence of Smad4. This indicates that complex formation with Smad4 is

not essential for dissociation from SARA or the receptor and suggests that phosphorylated R-Smads can form complexes lacking a Co-Smad (Fink et al., 2003). It seems likely that, in the formation of Smad heterotrimers, the two R-Smads activated by the two TβPI subunits of the receptor first form a dimeric complex with each other and that Smad4 is subsequently incorporated. How dimeric R-Smad–Co-Smad complexes might be assembled is not clear.

TLP is a cytoplasmic protein that associates with type II TGF-β receptors. It is associated with the receptor independently of TGF-β and appears to attenuate Smad3 signalling but not Smad2 signalling, possibly providing a mechanism to differentially regulate the extent to which these two proteins and their downstream responses are activated by TGF-β (Felici et al., 2003).

As mentioned above, TGF-β can signal via ALK1 as well as ALK5 (Oh et al., 2000). ALK1 is recruited into TGF-β receptor complexes in an ALK5-dependent manner and activates Smad1 and Smad5 in response to TGF-β. In this setting, Smad5 activation somehow inhibits the function of Smad2 and Smad3 at a point downstream of Smad2/3 phosphorylation. ALK1 expression levels are high in endothelial cells and it is thought that preferential activation of Smad1/5 signalling – rather than Smad2/3 signalling – is responsible for the mitogenic effect of TGF-β in this tissue (Fig. 1.4). This is consistent with the fact that ALK1 levels are low in epithelial cells and with the fact that BMP signalling via Smad1 and Smad5 stimulates proliferation by inducing *Id1* and other target genes (Goumans et al., 2003).

1.3.7 Smad nucleocytoplasmic shuttling

Before TGF-β induction Smad2, Smad3 and Smad4 are all continuously shuttling between cytoplasm and nucleus but the concentration of each is higher in the cytoplasm (Fig. 1.2) (Schmierer and Hill, 2005). Upon TGF-β induction, active Smad complexes accumulate in the nucleus, where they appear to remain for several hours. However, these complexes are continually being inactivated by an unknown process that involves dephosphorylation of R-Smads, complex disruption and recycling of Smad monomers to the cytoplasm. The recycled R-Smads can then be re-

phosphorylated by the receptor and incorporated again into active Smad complexes (Fig. 1.2). This mechanism is thought to allow continuous monitoring of receptor activity by nuclear Smad complexes (Inman et al., 2002; Nicolas et al., 2004; Pierreux et al., 2000).

Our understanding of the molecular mechanisms underlying the nucleocytoplasmic shuttling of Smad molecules remains limited. Smad2 and Smad3 have been reported to interact with the nucleoporins Nup153 and Nup214 independently of the nuclear transport receptors and independently of TGF-β, providing a possible mechanism of passive transport between cytoplasm and nucleus (Xu et al., 2003; Xu et al., 2002).

However, various lines of evidence indicate that active mechanisms may be more important in Smad nucleocytoplasmic transport. Smad3 and Smad4 each contain a canonical nuclear localisation signal (NLS) in their MH1 domain and this has been shown to mediate interaction with the Importin α/β nuclear transport system (Fig. 1.6) (Kurisaki et al., 2001; Xiao et al., 2003; Xiao et al., 2000a; Xiao et al., 2000b). This NLS is also present in Smad2 but, in the full-length version of this protein, its function appears to be masked by a 20-amino acid insert encoded by exon 3 (Fig. 1.6) (Kurisaki et al., 2001). Smad4 also contains a canonical nuclear export signal recognised by the export receptor CRM1, which mediates nuclear export of Smad4 (Fig. 1.6) (Pierreux et al., 2000; Watanabe et al., 2000). In both the presence and absence of TGF-β, Smad2 and Smad3 are exported independently of CRM1 but Smad3 is exported by Exportin 4 (Fig. 1.6) (Kurisaki et al., 2006) and the export of Smad2 is dependent on ATP, indicating that it too is exported by an active mechanism (Inman et al., 2002).

Whatever the precise mechanisms involved Smad2, Smad3 and Smad4 are capable of undergoing nucleocytoplasmic shuttling in both the presence and absence of TGF-β (Fig. 1.2). Upon TGF-β induction, the balance of this shuttling switches from cytoplasm to nucleus. Originally, it was proposed that this was the result of a release from cytoplasmic retention by SARA and its replacement with nuclear retention by DNA-bound transcription factors and other proteins (Xu et al., 2000). This model was supported by the fact that the H2 helix–β8/9 loop region of the MH2 domain of

Smad2 and Smad3 (Fig. 7) is involved in interactions with SARA, the nucleoporins and nuclear transcription factors (ten Dijke and Hill, 2004).

However, a recent study found no evidence for a change in the cytoplasmic mobility of Smad2 or Smad4 upon TGF-β induction, indicating that retention by SARA is not important for keeping Smad2 in the cytoplasm in the absence of TGF-β (Schmierer and Hill, 2005). On the other hand, the same study revealed that the nuclear mobility of Smad2 and Smad4 is lower in TGF-β induced cells, indicating that retention does indeed play a role in nuclear accumulation. This work also showed that the import rate constant for Smad2 remains unchanged by phosphorylation but that the export rate constant decreases. A combination of enhanced nuclear retention and reduced nuclear export – which may be the same thing – thus explains accumulation of Smads in the nucleus upon TGF-β stimulation (Schmierer and Hill, 2005).

1.3.8 Regulation of transcription by active Smad complexes

In the nucleus, active Smad complexes recognise the promoters of direct target genes by associating with cell-type specific transcription factors. These Smad-transcription factor complexes recruit coactivators and corepressors that enable the Smads to induce or repress the transcription of target genes (Fig. 1.8A) (Shi and Massague, 2003; ten Dijke and Hill, 2004). For more detail of Smad-mediated transcriptional control see sections 1.4 and 1.5 of this introduction.

1.3.9 Ending Smad signalling

A number of negative feedback mechanisms are responsible for shutting down the TGF-β pathway. Smad7 is a negative regulator of TGF-β receptor function that resides primarily in the nucleus in the absence of TGF-β and which accumulates in the cytoplasm upon ligand stimulation (Itoh et al., 1998). TGF-β also induces the transcription of *Smad7*, leading to enhanced pathway attenuation at later time points (Denissova et al., 2000). Smad7 functions by binding activated TβRI in the same

way as an R-Smad but it cannot be phosphorylated and instead acts to competitively inhibit the receptor and recruit factors that antagonise its function (Hayashi et al., 1997; Nakao et al., 1997). Smad7 associates with a protein phosphatase 1 (PP1) complex that has GADD34 as its regulatory subunit and with the Smad-ubiquitin regulatory factors Smurf1 and Smurf2. PP1 is thus able to dephosphorylate and inactivate TβRI while the Smurfs act as E3 ubiquitin ligases to mediate polyubiquitination of both receptor types and subsequent degradation by the proteasome (Ebisawa et al., 2001; Kavsak et al., 2000; Shi et al., 2004). This process of receptor degradation involves internalisation via lipid rafts and caveolae (Di Guglielmo et al., 2003). In the case of acute signalling, these mechanisms lead to complete shutdown of the pathway within nine hours of induction (Inman et al., 2002). Smad6 is also a negative regulator of the pathway (Imamura et al., 1997) and is thought to act in a similar fashion to Smad7 but its role has not been extensively studied.

1.3.10 Other mechanisms for antagonising Smad function

Cells possess a range of other mechanisms that can be used to antagonise Smad function. As already described, ERK-mediated phosphorylation in the linker region may be able to inhibit Smad function (Kretzschmar et al., 1999). Similarly, in epithelial cells and fibroblasts, Cdk2 and Cdk4 can phosphorylate sites in the linker region of Smad3, resulting in attenuated induction of $p15^{INK4B}$ and reduced repression of *c-Myc* (Matsuura et al., 2004). Calcium-calmodulin-dependent protein kinase II (CaMKII) – which is also activated downstream of growth factor signalling – similarly antagonises Smad function by phosphorylating several sites in the linker region (Wicks et al., 2000). It is likely that all these mechanisms function to suppress the cytostatic response to TGF-β in cells that are already committed to G_1 progression.

A similar function may be achieved by interaction of Smad3 with the anti-apoptotic kinase PKB/Akt. This interaction has been reported to take place when PKB/Akt is activated downstream of PI3K and seems to be independent of the kinase activity of

Akt. Functionally, the effect is to sequester Smad3 in the cytoplasm. This appears to attenuate the apoptotic but not the cytostatic response to TGF-β, suggesting that the cytostatic response requires lower levels of signal (Conery et al., 2004).

It has also been shown that TGF-β can induce the expression of growth factors such as TGF-α and PDGF and can activate ERK and JNK in a strong but indirect fashion (see below). As with the induction of Smad7, all these effects may act as negative feedback mechanisms to shut down the TGF-β pathway (Gotzmann et al., 2006; Vinals and Pouyssegur, 2001). Indeed mitogen signalling has also been shown to induce *Smad7*, supporting the idea that indirect activation of MAP kinases is an important negative feedback mechanism in TGF-β signalling (Uchida et al., 2001).

1.3.11 Smad-independent pathways of TGF-β signalling

It has been shown that, as well as activating the Smad pathway, TGF-β can also signal through pathways that do not involve the Smads. These include true Smad-independent pathways that are activated directly at TGF-β receptor complexes and pathways that are activated indirectly via Smad-dependent gene responses. ERK, JNK and p38 MAPK all seem to fall into both categories but the direct responses observed are very weak and high levels of activation are observed only when indirect responses take effect (Derynck and Zhang, 2003).

Direct activation of ERK, JNK and p38 MAPK by TGF-β receptors seems to involve activation of both Ras and TGF-β-activated kinase 1 (TAK1) – in the latter case by a mechanism that may involve XIAP as an adaptor (Yamaguchi et al., 1999; Yamaguchi et al., 1995; Yue and Mulder, 2000). Like Raf, TAK1 is a MAP kinase kinase kinase (MAPKKK) family member and upon TGF-β stimulation may activate JNK via MEK3 or MEK6 and p38 MAPK via MEK4 (Yamaguchi et al., 1995). Similarly, activation of Ras by TGF-β is thought to lead to activation of ERK via MEK1 or MEK2 (Yue and Mulder, 2000).

Indirect activation of ERK by TGF-β seems to occur by transcriptional induction of *TGF-α*, *PDGF* and other growth factors. It has been shown that endothelial cells secrete TGF-α when treated with TGF-β and that this binds to ErbB1 (a receptor tyrosine kinase that also binds EGF) and activates ERK1 and ERK2. Treatment of cells with soluble ErbB1 or an inhibitor of EGF receptor kinase activity blocked the delayed activation of ERK1 and ERK2 by TGF-β (Vinals and Pouyssegur, 2001). Similarly, it has been shown that hepatocytes produce bioactive PDGF-A and both subunits of the PDGF receptor in response to TGF-β (Gotzmann et al., 2006).

Interestingly, HGF and other mitogens also induce the expression of TGF-β, indicating the existence of a complex paracrine network controlling the homeostasis of mammalian tissues (Wick et al., 2001). Furthermore, it has been shown that SB-431542, an inhibitor of TβRI kinase activity, inhibits hepatocellular EMT induced by HGF (Levy and Hill, 2005) and that dominant-negative PDGF receptors inhibit hepatocellular EMT induced by TGF-β (Gotzmann et al., 2006). This suggests that TGF-β and growth factors can cross-activate each other and that EMT takes place only when both pathways are active. Another functional role of TGF-β-dependent activation of ERK is in angiogenesis, where interference with ERK impairs the ability of TGF-β to induce capillary formation (Vinals and Pouyssegur, 2001).

Indirect activation of p38 MAPK seems to rely on transcriptional activation of *GADD45β*. This has been shown to rely on functional Smad molecules and antisense inhibition of GADD45β expression abrogates delayed activation of p38 MAPK by TGF-β (Takekawa et al., 2002). Similarly, indirect activation of JNK by TGF-β may rely on transcriptional induction of JKAP, a dual-specificity phosphatase that acts at an unknown point upstream of JNK (Chen et al., 2002a). The functional significance of JNK and p38 MAPK in TGF-β signalling unclear but it is known that JNK activation can act as a negative feedback mechanism controlling Smad function (see above).

Another well-characterised pathway activated by TGF-β operates through the scaffolding protein Par6 and, unlike the Smad-independent pathways already described, Par6-mediated TGF-β-signalling appears to operate in a direct mode only.

Par6 interacts with TβRI independently of TGF-β and is phosphorylated by TβRII in response to TGF-β stimulation. This contributes to TGF-β-induced EMT by enabling Par6 to interact with Smurf1. Smurf1 then polyubiqitinates RhoA, leading to its degradation and the resultant dissolution of tight junctions and a loss of cell polarity. Interestingly, this mechanism may clear RhoA only from the lateral submembrane compartment of the cell and is accompanied by just a 30% reduction in cellular levels of RhoA (Ozdamar et al., 2005).

The hypothesis that TGF-β-induced RhoA degradation is localised may be important because RhoA and the related small GTPases Rac and Cdc42 are also activated by TGF-β. These activations occur by both direct and indirect mechanisms and have been implicated in TGF-β-induced EMT and cell-cycle arrest (Bakin et al., 2002; Edlund et al., 2002; Engel et al., 1998). The mechanisms through which TGF-β directly activates small GTPases have not been elucidated but indirect activation of RhoA has been shown to occur by Smad-dependent transcriptional activation of *NET1*. NET1 is a guanine nucleotide exchange factor (GEF) that stimulates RhoA to exchange GDP for GTP. Functionally, treatment of cells with an inhibitor of ROCK (a kinase activated directly by RhoA) blocks the TGF-β-induced stress fiber formation and impairs TGF-β-induced EMT (Shen et al., 2001). Taken together with the data obtained from studies with Par6, these results indicate that RhoA has a complex role in TGF-β-induced EMT.

TGF-β can also activate phosphatidylinositol-3-kinase (PI3K) and, as before, this seems to occur by a mix of direct and indirect mechanisms (Derynck and Zhang, 2003). Direct activation may operate via RhoA and it has been shown that indirect activation depends on TGF-α (Vinals and Pouyssegur, 2001). As with ERK, inhibition of PI3K activity impairs TGF-β-induced angiogenesis, demonstrating that activation of other signalling pathways is required for TGF-β function. Importantly, some PI3K activity seems to be needed for TGF-β–Smad signalling (Bakin et al., 2000). This is probably due to a requirement by SARA for PI(3)P in the plasma membrane and is another reflection of the complex paracrine network that operates in mammalian tissues.

Finally, TGF-β appears to directly activate a form of protein phosphatase 2A (PP2A) that dephosphorylates and inactivates $p70^{S6K}$. This seems to occur by sequestration of the PP2A subunit Bα at active TGF-β receptors. When active, $p70^{S6K}$ enhances general protein synthesis by phosphorylating and activating the ribosomal S6 protein. By inhibiting this kinase, TGF-β may suppress growth in cell size concomitant with its effect on proliferation (Petritsch et al., 2000).

1.4 Smad–transcription factor interactions

Treatment of cells with TGF-β results in up- or down- regulation of several hundred direct and indirect TGF-β target genes (Levy and Hill, 2005). Direct target genes are controlled by binding of active Smad complexes to sites in their promoter region in cooperation with other transcription factors (Fig. 1.8A). These Smad partners are required to confer binding strength and promoter selectivity and it is thought that availability of cell-type specific transcription factors dictates different TGF-β responses (Massague et al., 2005). A number of themes are evident in Smad-transcription factor cooperation (see below). Examples of such themes include: a) the use of conserved sequence motifs through which transcription factors interact with Smad molecules, b) self-enabling and self-inhibiting responses where direct Smad target genes encode transcription factors that go on to promote or repress Smad function; c) interaction with ubiquitous factors whose function may be regulated at levels other than gene expression; d) interaction with factors that are controlled by other extracellular signals; e) interaction of Smads with transcription factors that – like the Smads – also function as tumour suppressors or tumour promoters and f) deregulation of Smad transcriptional partners in tumours.

1.4.1 DNA binding by Smad complexes

In the nucleus, active Smad complexes bind directly to the Smad-binding element (SBE) found in the promoters of direct TGF-β target genes. The canonical SBE – as identified by site-selection studies – consists minimally of the sequence AGAC but

multiple AGAC repeats are more common (Fig. 1.9A) (Johnson et al., 1999; Yingling et al., 1997; Zawel et al., 1998). Smad-DNA binding occurs directly via the β-hairpin motif found in the MH1 domain of Smad3 and Smad4. This motif consists of two short anti-parallel β strands joined by a linker loop and fits into the major groove over the SBE (Fig. 1.7A). It represents a novel DNA binding structure and makes sequence-specific hydrogen-bond contacts with three of the four base-pairs in an AGAC Smad-binding site. This structure is conserved in all DNA-binding Smad molecules and, as such, cannot be responsible for promoter discrimination by different Smad complexes. Smad MH1 domains do not appear to interact with each other either in solution or in the context of DNA-bound complexes (Shi et al., 1998).

Full-length Smad2 is unable to bind the SBE directly due to the 20 amino-acid insert encoded by exon 3, which is located adjacent to the β-hairpin (Fig. 1.6). However, a minor isoform lacking this insert is able to bind DNA (Yagi et al., 1999). This splice variant is expressed in different tissues at levels ranging from one-tenth to one-third that of the full-length protein. Mice lacking Smad2 die at gastrulation (Heyer et al., 1999; Waldrip et al., 1998) but mice engineered to express only the Smad2ΔExon3 isoform are viable and fertile, indicating that – whatever its actual role – exon 3 is dispensable for most functions of Smad2 (Dunn et al., 2005).

Some TGF-β responsive regions lack a canonical SBE, suggesting either that some Smad responses are independent of direct Smad-DNA binding or that Smads can bind DNA sites other than the canonical SBE. This second possibility is supported by the fact that a certain degree of "wobble" is permitted in the sequence of the SBE. The crystal structure of the Smad MH1 comain in complex with DNA shows that the third base-pair of the AGAC element does not make direct contacts with the β hairpin (Shi et al., 1998) and in the Smad-binding region of the *c-Myc* promoter, Smad3 or Smad4 seems to binds an AGCC site instead (Chen et al., 2002b).

Smads may also bind more divergent DNA sequences and various GC-rich motifs have been proposed as degenerate Smad-binding sites. *Drosophila* Medea (the Smad4 homolog) has been shown to bind a GCCGNCGC sequence in the promoters of *Vestigal* and *Tinman* (Kim et al., 1997; Xu et al., 1998) and the BMP-responsive

element in the *Smad6* promoter contains four overlapping copies of a GC-rich motif (Ishida et al., 2000). Interestingly, the *Id1* promoter contains a GC-rich element along with a canonical SBE and both elements are required for BMP-mediated induction via Smad1-Smad4 complexes and TGF-β-mediated repression via Smad3-Smad4 complexes, the difference being determined by the recruitment of ATF3 to Smad3-Smad4 complexes (Kang et al., 2003; Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al., 2002). This and other findings, along with the divergence in sequence between the SBE and GC-rich elements, has led to speculation that the MH1 domain contains a second DNA-binding motif that can contact various GC-rich motifs (Shi and Massague, 2003). One candidate is the positively charged H2 helix, which lies adjacent to the β-hairpin in the tertiary structure of Smad3 and Smad4 (Fig. 1.7A) (Shi et al., 1998). It has been documented that clusters of lysine – such are present in this helix – preferentially bind GC-rich sequences (Choo and Klug, 1997), making the H2 helix a strong candidate for the motif that recognises GC-rich Smad-binding sites.

The stoichiometry of SBE-bound Smad complexes has been a matter of some controversy – as has been that of inactive Smads and active Smad complexes in solution. It now seems that heterotrimers composed of one Co-Smad and two R-Smads predominate, both in solution and on DNA (Massague et al., 2005). However, multi-epitope tagging followed by electrophoretic mobility shift assay with antibody supershifts reveals that Co-Smad–R-Smad dimers are also important on certain promoters (Fig. 1.9B) (Inman and Hill, 2002). Since no crystal structure of a Smad dimer or trimer bound to a DNA sequence derived from a natural promoter has yet been obtained, we do not know the detailed physical architecture adopted by such complexes. Nor do we know the precise physical arrangements that accommodate Smad-transcription factor interactions on DNA (Massague et al., 2005).

1.4.2 Interaction of Smads with transcription factors

To achieve promoter recognition, Smad complexes require the cooperation of interacting transcription factors that bind sites adjacent to the SBE (Massagué and Wotton, 2000). Associated transcription factors are essential since the affinity of

Smads for the SBE (around 10^{-7} M) is insufficient to give independent DNA binding at physiological concentrations (Shi et al., 1998). In addition, Smad complexes may be incapable of recruiting the basal transcription apparatus without the help of associated transcription factors (Ross, S.J. *et al*, manuscript submitted). If it were not for these facts, juxtaposition of an AGAC triple-repeat with a TATA box would, in principle, be sufficient to confer promoter specificity.

In addition, Smad–transcription factor interactions function as a major determinant of cellular responses to TGF-β. It is thought that different Smad-interacting transcription factors are expressed in different cell types and that this enables TGF-β to induce different sets of target genes in different cell types (Shi and Massague, 2003). This is illustrated schematically in Figure 1.8A where an hypothetical Smad partner, transcription factor A, is expressed in cell type A and another hypothetical partner, transcription factor B, is expressed in cell type B. Active Smad complexes require factor A to induce gene 1 and factor B to induce gene 2. The result is that gene 1 is induced by TGF-β in cell type A but not cell type B and gene 2 is induced by TGF-β in cell type B but not cell type A.

Smad target genes can contain just a single SBE (usually a CAGAC sequence) closely juxtaposed with a binding site for a Smad-interacting transcription factor. The best example of this is the *Xenopus Mix.2* promoter, where an SBE is juxtaposed to a FoxH1 binding site (Chen et al., 1996). When the partner has strong affinity and specificity for its cognate binding site, one SBE is sufficient to allow assembly of a Smad-transcription factor-DNA complex and TGF-β-dependent transcriptional control. At least some Smad-interacting transcription factors, such as FoxH1b, are pre-bound to DNA in the promoters of TGF-β target genes (Howell et al., 2002) but recruitment of active Smad complexes clearly results in substantial changes in the level of transcription. This must be due to recruitment of additional coregulators or more efficient recruitment of coregulators than is possible in the absence of the Smads. Other transcription factor partners, such as E2F4/5, associate with Smads in solution before binding DNA as a pre-formed complex (Chen et al., 2002b). There are also complexes where Smads and (as yet unidentified) partners bind DNA simultaneously without previously having been associated in solution and without

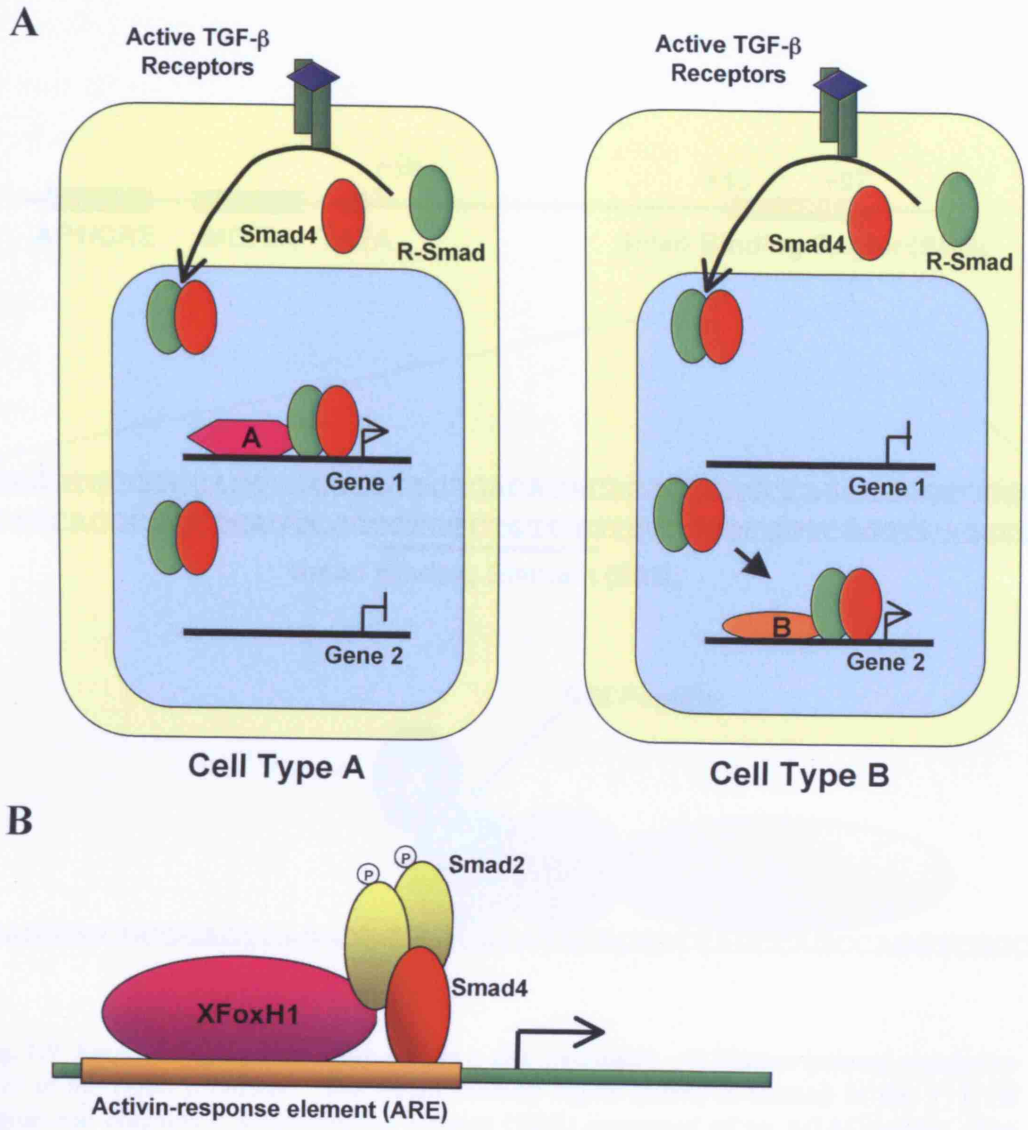


Fig. 1.8 Smad-transcription factor interactions. *A) Cooperation of Smads with cell-type-specific transcription factors: a general model.* Cell type A expresses transcription factor A and cell type B expresses transcription factor B. Smads require factor A to induced gene 1 and, as a consequence, gene 1 is induced by TGF-β in cell type A, but not cell type B. Conversely, Smads require factor B to activate gene 2 and this gene is induced in cell type B but not cell type A .*B) Interaction of Smad2 and Smad4 with XFoxH1 on the activin-response element (ARE).* One Smad2 molecule contacts FoxH1 through the SIM and the other through the FM. Smad4 does not make direct contact with FoxH1. Smad4 and FoxH1 both contact DNA directly but specificity and affinity are conferred mainly by FoxH1. When it includes FoxH1a this complex is known as ARF1 and when it includes FoxH1b it is known as ARF2.

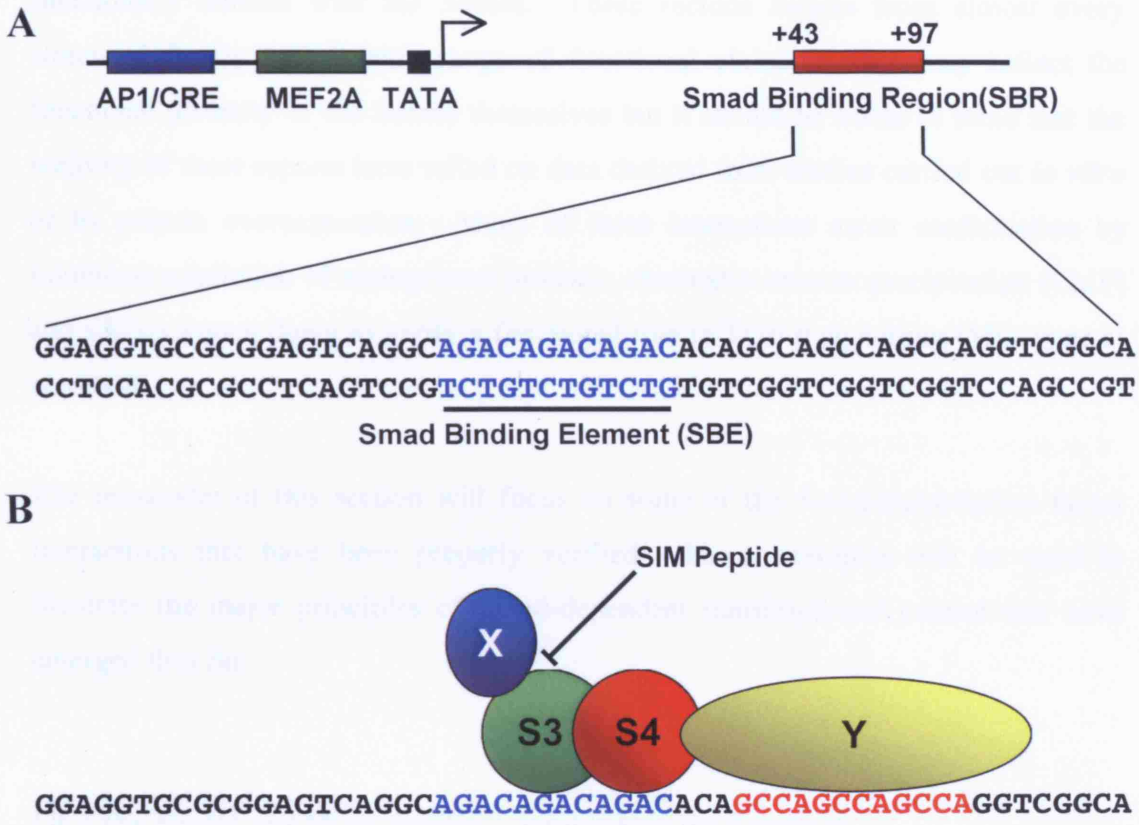


Fig. 1.9 Smad complex formation on the *c-Jun* promoter. *A) Transcriptional regulatory sites in the* c-Jun *promoter.* The Smad binding region (SBR) is located in the 5' UTR region and contains a Smad-binding element (SBE) composed of an AGAC triplet. The SBR cooperates with c-Jun binding to an upstream AP1/CRE site to induce transcription in a TGF-β-dependent manner. The MEF2A binding site is also shown. *B) Smad complex formation on the* c-Jun *SBR.* A Smad3-Smad4 heterodimer is bound to the SBE. This dimer is associated with a transcription factor (factor Y) bound to a downstream GCCA triplet and Smad3 is directly bound by a SIM-containing coregulator (factor X). The Smads and factor Y depend on each other for DNA binding and factor X makes a major contribution to complex stability.

either Smads or partners having been pre-bound to DNA (see section 1.6 below) (Inman and Hill, 2002).

At least 60 regulatory transcription factors have been reported to physically and/or functionally interact with the Smads. These include factors from almost every structural family and a wide range of functional classes. This may reflect the functional diversity of the Smads themselves but it should be borne in mind that the majority of these reports have relied on data derived from studies carried out *in vitro* or by protein overexpression. Many of these interactions await confirmation by immunoprecipitation of endogenous proteins, chromatin immunoprecipitation (ChIP) and siRNA knock down to verify a functional role in TGF-β signalling (Massague et al., 2005).

The remainder of this section will focus on some of the Smad-transcription factor interactions that have been properly verified. These examples will be used to illustrate the major principles of Smad-dependent transcriptional control that have emerged thus far.

1.4.3 The FoxH1 and Mixer families – establishing the paradigm

Early *Xenopus* development is the setting where Smad-transcription factor interactions have been best characterised. This is particularly true for members of the FoxH1 and Mixer families. These factors have been shown to undergo a direct Activin/Nodal-dependent interaction with Smad2 in solution and on DNA. They are also required for Activin/Nodal gene responses and for these signals to fulfil their proper developmental functions. They thus provide a model for understanding Smad-transcription factor interactions in general.

The first Smad transcriptional partner identified in *Xenopus* (or any organism) was FoxH1a, originally known as FAST-1 (for forkhead activin signal transducer) (Chen et al., 1996). FoxH1a contains an N-terminal forkhead/winged helix DNA binding domain. In this domain, a helix-turn-helix motif very similar to that found in

homeodomian factors is accompanied by a β sheet structure containing two loops that project from the protein as wings. Forkhead proteins can bind DNA through either the helix-turn-helix motif or the winged motif (Gajiwala and Burley, 2000).

Electrophoretic mobility shift assay (EMSA) with Activin-injected *Xenopus* embryos revealed the existence of Activin-inducible factor 1 (ARF1), a complex that binds the 50 bp Activin-response element (ARE) of the *Mix.2* promoter. Unlike ARF2, which forms at later stages in response to endogenous Activin/Nodal signalling (see below), high levels of ARF1 are formed only in ligand-injected embryos. Screening of a cDNA expression library identified FoxH1a as a novel protein present in this complex and antibody supershifts show that it is associated with Smad2 (Chen et al., 1996).

Subsequent work demonstrated that Smad4 is also a component of ARF1 and that, in solution, both Smad2 and Smad4 associate with FoxH1a in an Activin-dependent fashion. The site of interaction of Smad2 with FoxH1a was mapped to the MH2 domain of Smad2 and to a region near the C-terminus of FoxH1a that was dubbed the Smad-interaction domain (SID). In ARF1 complexes, both Smad4 and FoxH1a make direct contacts with DNA (Fig. 1.8B) but only the contacts made by FoxH1a are essential for DNA binding and gene regulation – although mutagenesis of the SBE does weaken the complex. Smad4 seems to bind a single CAGAC site and an adjacent GC-rich sequence while FoxH1a contacts an ATGTGTATTC site located 8 bp upstream (Chen et al., 1997).

From early gastrulation, a much more abundant ARF complex, known as ARF2, is formed on the *Mix.2* ARE in response to endogenous TGF-β superfamily signalling (Howell et al., 2002). This complex contains FoxH1b/Fast3, a paralog of FoxH1a. The N-terminal forkhead domain is highly conserved, as are parts of the SID (specifically the SIM and FM motifs – see below) but intervening regions are more divergent. FoxH1b is also much smaller than FoxH1a. Even in Activin-injected embryos, ARF1 complexes disappear as gastrulation progresses and are replaced by ARF2, reflecting a greater affinity of FoxH1b for the FoxH1 binding site (Howell et al., 2002). It should be noted that, prior to mid-gastrulation, *Xenopus* embryos express Smad4β instead of Smad4α and this is the Co-Smad component of

endogenous ARF complexes. Smad4β is encoded by a distinct gene from Smad4α and, while Smad4α is very similar to human Smad4, Smad4β is more divergent (Howell et al., 1999).

FoxH1a is expressed in *Xenopus* from the oocyte until shortly after gastrulation, at which time its expression levels decline (Chen et al., 1996). In early gastrulation it is localised to the animal pole and prospective mesoderm in the marginal zone (Howell et al., 2002). Overexpression of this factor results in enhanced expression of mesoendodermal markers (Xnr target genes) and induction of a partial secondary axis (Watanabe and Whitman, 1999). Morpholino knock-down of FoxH1a results in slightly lowered expression of *Xlim-1*, *XFKH1* and *Xbra* but not of other mesoendodermal markers. However, gastrulation movements are disrupted in these animals, blastopore closure is delayed, the primary axis is severely shortened and head structures are lost (Howell et al., 2002). These findings indicate that FoxH1a cooperates with Activin/Nodal signalling to induce gastrulation and establish of the primary body axis.

FoxH1b is expressed in a narrow time window during early- to mid- gastrulation and is then rapidly downregulated. Like FoxH1a, it localises to the animal pole and prospective mesoderm in the marginal zone. However, ARF2 differs from ARF1 in that it is strongly induced by Derriere. Activin enhances ARF2 formation to a lesser extent and Xnr has no effect, indicating that endogenous Xnrs probably signal via ARF1 whereas endogenous Activin and Derriere signal via the more abundant ARF2 complex. Functionally, FoxH1b seems to play a similar role to FoxH1a. The knockdown phenotype is like that of FoxH1a but is less severe, probably reflecting the fact that FoxH1a is expressed earlier. Simultaneous interference with the function of both factors produces a very severe phenotype involving virtually complete loss of axis formation, failure to close the neural tube and disintegration due to a loss of epithelial integrity. These findings reflect an absolute requirement for the FoxH1 family in Activin/Nodal family signalling in the early *Xenopus* embryo (Howell et al., 2002).

A single FoxH1 family member is also found in mammals. Due to sequence differences, this protein was originally referred to as Fast1 in humans and Fast2 in mouse, but both are now simply referred to as FoxH1 (Labbe et al., 1998; Whitman, 2001; Zhou et al., 1998). *FoxH1*$^{-/-}$ mouse embryos show phenotypes of varying severity involving impaired axis formation and loss of anterior structures – again reflecting a genetic interaction with Activin/Nodal signalling (Hoodless et al., 2001; Yamamoto et al., 2001). FoxH1 is also involved in induction of *Lefty2* in response to Nodal and establishment of left-right asymmetry (Saijoh et al., 2000).

Further well-understood Smad-interacting transcription factors in early embryos are members of the Mixer family. This is a group of factors that are thought to bind DNA through both a paired-like domain and a homeodomain. It includes Mix.1, Mix.2, Mixer/Mix.3, Mix.4/Bix1, Milk/Bix2, Bix3 and Bix4 (Germain et al., 2000).

During *Xenopus* gastrulation, the *goosecoid* promoter is induced by Wnt signalling to the proximal element (PE) and Activin/Nodal signalling to the distal element (DE) (Watabe et al., 1995). EMSA shows that Mixer and Milk bind the DE in a complex that also contains Smad2 and Smad4. In cultured cells, Mixer was shown to bind Smad2 independently of TGF-β and to recruit Smad4 in a TGF-β-dependent manner. This is in accordance with reporter assays showing that Mixer or Milk is required for TGF-β-induced transcription from the DE (Germain et al., 2000).

Functionally, Mixer and Milk are induced by Nodal signalling during gastrulation (see below) and localise to the marginal zone. It is likely that they are involved in induction of mesoendodermal genes and gastrulation movements by Activin/Nodal signalling (Randall et al., 2002). Importantly, it has been shown that Zebrafish double knockouts of Bon (the Mixer homolog) and Sur (the FoxH1 homolog) have a phenotype that corresponds much more closely to that resulting from complete loss of Nodal signalling than either mutant alone. Bix3 is the only other member of the Mix family interact with Smad2 (Kunwar et al., 2003). Misexpression of Bix3 in *Xenopus* upregulates mesoendodermal markers, suggesting that it also plays a role Activin/Nodal signalling during gastrulation (Trindade et al., 2003).

Taken together, these findings provide the basis for the general model of Smad–transcription factor interactions described above and illustrated in figure 1.8A. Smads interact physically with members of the FoxH1 and Mixer families and these complexes associate with the promoters of TGF-β superfamily target genes expressed in early embryos. Smads are dependent on these partners to induce the relevant target genes and the phenotypes induced by depletion of FoxH1 or Mixer family members resemble those induced by interference with Activin or Nodal signalling.

1.4.4 The Smad Interaction Motif (SIM) and Fast Motif (FM)

As well as embodying the basic principles of Smad-transcription factor interactions, members of the FoxH1 and Mixer families are important because the physical basis of their interaction with Smad molecules have been very well characterised. Both families contain conserved sequence motifs that mediate direct interactions with defined pockets on the surface of Smad2 and Smad3 (Fig. 1.7B) (Germain et al., 2000; Randall et al., 2002; Randall et al., 2004). One of these motifs is functionally similar to the Smad-binding domain (SBD) of SARA and it is thought that functionally equivalent motifs will be found in other Smad-interacting factors (Randall et al., 2002).

Sequence alignment between FoxH1a, Mixer and Milk led to the identification a conserved Smad interaction motif (SIM), characterised by a core PPNK sequence (Germain et al., 2000). This motif binds to a conserved hydrophobic pocket adjacent to the H2 helix and β8/9 loops on the MH2 domain of Smad2 and Smad3 (Fig. 1.7B) (Randall et al., 2002). Only the interaction with Smad2 is important in gastrulating embryos, since Smad3 is not expressed at this stage (Howell et al., 2001). The SIM is also present in mammalian FoxH1 and in zebrafish Bon and Sur (Howell, M and Hill, CS, unpublished data). The only other factors where this motif is known to be conserved are *Xenopus* FoxH1b and Bix.3 (Germain et al., 2000; Howell et al., 2002), however there is evidence that cells derived from adult human epidermis – where none of these Smad partners are expressed – contain factors with a sequence functionally equivalent to the SIM (Inman and Hill, 2002). In the FoxH1 and Mixer

families, the SIM is located towards the C-terminus, far from the N-terminal DNA-binding domains. The Mixer SIM is both necessary and sufficient to mediate TGF-β-dependent transcription, as demonstrated by experiments in which it was fused with the DNA-binding domain of Gal4. Mix.1 lacks a SIM and insertion of the Milk SIM into this factor is sufficient to convert it into a strong mediator of TGF-β-dependent transcription (Germain et al., 2000).

Extensive mutational analysis revealed that a functional SIM requires residues C-terminal to the core PPNK motif (which in located between residues 291 and 294 in Mixer) and showed that the minimal SIM is 15 residues in length. As mentioned above, a similar sequence is found in the Smad-binding domain (SBD) of SARA, providing a putative basis for *in vivo* competition between SARA and nuclear transcription factors. When the SIM was first identified, the co-crystal structure of Smad2 with SARA had already been solved, providing a basis for molecular modelling and structure-directed mutagenesis of the Smad2-SIM interface (Fig. 1.7B) (Randall et al., 2002).

This modelling identified the binding site for the Mixer SIM as a shallow hydrophobic cleft on the surface of the Smad2 MH2 domain, adjacent to the H2 helix and the β8/9 loops (Fig. 1.7B). In this model, the interaction of Smad2 with the SIM is mediated by a combination of hydrophobic interactions and hydrogen bonds. Important hydrogen bonds bridge P292 of the SIM with W368 and T297 of Smad2. Important hydrophobic interactions links P292 of the SIM with W368 and Y366 of Smad2. Further key hydrophobic interactions link M300 of the SIM with W368 and C374 of Smad2. This analysis identified Smad2 residue W368 as perhaps the most critical residue involved in interaction with the SIM (Fig. 1.7B) (Randall et al., 2002).

Mutagenesis studies confirmed the importance of P292 and T297 in the SIM and W368 in Smad2, along with other residues involved in the Smad2-SIM interaction. They also confirmed the importance of N293, which was predicted by the model to be central in a network of hydrogen bonds that give the SIM a rigid coil conformation similar to that adopted by the SARA SBD (Fig. 1.7B). Most of the residues important

in the Smad2-SIM interaction are absent from Smad1, making it likely that SIM-containing factors do not play a role in BMP signalling (Randall et al., 2002).

Xenopus FoxH1a and FoxH1b also contain a Fast motif (FM) located some distance upstream of the SIM. This is also present in their mammalian and zebrafish homologs but is absent from the Mixer family (Germain et al., 2000). Together, the FM and the SIM make up the SID of FoxH1 factors (Howell et al., 2002) and both elements are conserved in mammalian FoxH1. The 16 amino acid FM is characterised by two regions within it that are conserved within the FoxH1 family. These are FM1, a 5-residue sub-motif with the consensus sequence LPTSY and FM2, a 6-residue sub-motif with the more variable consensus sequence PN(V/A)V(A/M)P(L/P). Molecular modelling indicates that the FM binds a hydrophobic pocket in Smad2 overlapping that bound by the SIM and including Smad2 residues W368 and V373. However, the FM binding site extends much closer to the Smad2-Smad4 interface than the SIM-binding pocket, with FM2 making contact with V341 near the interface (Randall et al., 2004).

The SIM, FM1 and FM2 motifs are all required for ARF1 complex formation but FM1 is dispensable for transcriptional activation, reflecting the fact that suboptimal DNA binding can be sufficient for full reporter activation (Randall et al., 2004). ARF2 assembly on the other hand relies only on the FM2 – mutation of the SIM and FM1 having much lesser effects on complex formation. This reflects the greater stability of ARF2. Likewise, for transcriptional activation of an ARE-driven reporter, only FM2 is required (Howell et al., 2002).

Multi-epitope tagging and EMSA-supershift experiments reveal that ARF1 contains two Smad2 molecules, one Smad4 and one FoxH1a (Fig. 1.8B) (Inman and Hill, 2002). It is thought that the SIM mediates interation of FoxH1a with one Smad2 subunit while the FM mediates interaction with the other (ten Dijke and Hill, 2004). Similarly, the FM may mediate preferential incorporation of Smad2 even when Smad3 is overexpressed.

1.4.5 Self-enabling and self-inhibiting Smad responses

Many of the genes induced by Smads include transcription factors such as Slug, Snail, and TIEG1 (see sections 1.2.4 and 1.2.5 above) (Levy and Hill, 2005). These function to mediate delayed cellular responses to TGF-β. Another function of transcription factors upregulated in response to TGF-β can be to modulate Smad gene responses. This includes self-enabling responses in which TGF-β-induced transcription factors associate with active Smad complexes to promote their function and self-inhibiting responses in which they have the opposite effect. This form of regulation may provide a platform for other signals to modulate the TGF-β response or may function to make certain indirect gene responses dependent on sustained TGF-β signalling. Two groups of transcription factors that participate in TGF-β responses of this kind include the AP1 and Mixer families (Massague et al., 2005).

In *Xenopus*, *Mixer* is induced by Activin/Nodal signalling at gastrulation (see section 1.4.3 above) and associates with active Smad complexes on the promoters of *goosecoid*, *Sox17* and other Nodal target genes. This allows induction of these genes and is essential for normal germ-layer specification and proper gastrulation movements (Germain et al., 2000). These findings represent the first well-characterised example of a Smad-induced transcription factor cooperating with Smads in other gene responses.

The AP-1 (activating protein 1) group of bZIP transcription factors encompasses the Jun, Fos, Fra and ATF subfamilies. These are widely expressed, can bind DNA as homo- or hetero-dimers, can associate with AP-1 and CRE sites and are involved in a wide range of biological processes. *c-Jun* and *c-Fos* are both induced downstream of ERK and participate in the early events of G_1 progression. Similarly, c-Jun, JunB, JunD and ATF2 are phosphorylated and activated by JNK downstream of stress signals (Hess et al., 2004).

Smad and AP-1 function are thought to intersect at several levels. *c-Jun*, *JunB*, *c-Fos* and *ATF3* have all been shown to be induced by TGF-β signalling (Kang et al., 2003; Li et al., 1990; Pertovaara et al., 1989; Verrecchia et al., 2001a) and the protein

products of all these plus JunD are each reported to interact physically and/or functionally with Smads (Dennler et al., 2000; Kang et al., 2003; Liberati et al., 1999; Pessah et al., 2002; Verrecchia et al., 2001b). These findings are the basis for the view that AP-1 factors function in self-enabling and self-inhibiting responses to TGF-β. In addition to this, Smad and TGF-β function are likely to be directly and indirectly modulated by Jun N-terminal kinase (JNK) (Pardoux and Derynck, 2004; Ventura et al., 2004).

Many reports have shown that Smad2 and Smad3 can interact with overexpressed c-Jun, JunB or JunD independently of TGF-β (Dennler et al., 2000; Liberati et al., 1999; Pessah et al., 2002; Verrecchia et al., 2001b). However, these findings have not been confirmed with endogenous proteins. Nevertheless, it is very likely that Smad proteins do physically interact with members of the Jun and Fos subfamilies, at least in the context of DNA-bound complexes. One line of evidence that supports this conclusion is EMSA analysis of the *COL7A1* promoter in which a Smad-containing protein-DNA complex is supershifted by pan-Jun and pan-Fos antibodies (Verrecchia et al., 2001b). Similarly, a TGF-β-dependent nucleoprotein complex forms on AP-1 consensus oligonucleotides and is supershifted by antibodies specific to Smad3, Smad4, c-Jun and c-Fos (Zhang et al., 1998). Only in the case of ATF3 is there evidence that endogenous protein interacts with endogenous Smads in solution. This interaction seems to be independent of TGF-β (Kang et al., 2003).

One set of findings shows that Smad and AP-1 family members can cooperate in transcriptional induction, providing the basis for self-enabling responses (Kang et al., 2003; Qing et al., 2000; Wong et al., 1999; Zhang et al., 1998). In reporter assays, consensus AP-1 sites have been found to mediate transcriptional activation in a TGF-β-dependent manner and this seems to be potentiated by co-transfection with Smad3 and Smad4 (Zhang et al., 1998). Similar studies suggest that Smad3-Smad4 complexes cooperate with c-Jun on the *Collagenase I* and *c-Jun* promoters (Qing et al., 2000; Wong et al., 1999). Likewise, chromatin immunoprecipitation reveals that Smad3, Smad4 and ATF3 are recruited to the *Id1* promoter in a TGF-β-dependent manner, leading to cooperative repression of *Id1* and cell cycle arrest (Kang et al., 2003). The *Collagenase I* and *Id1* promoters both contain AP-1 sites closely

juxtaposed with SBEs (Kang et al., 2003; Qing et al., 2000) while the *c-Jun* promoter contains widely separated AP-1 and SBE sites that cooperate functionally (Wong et al., 1999).

However, other work has found that AP-1 factors can also antagonise Smad function, indicating that the functional relationship of Smads with AP1 is promoter-specific and providing the basis for self-inhibiting responses. Transfection with c-Jun or JunB inhibits TGF-β-dependent activation of reporters driven by a multimerised SBE (Dennler et al., 2000). In agreement with this, expression of antisense *c-Jun* or *JunB* can slightly potentiate reporter activation (Verrecchia et al., 2001a). These effects must occur via functional interaction with Smad3-Smad4 complexes since Smad2 is not recruited to these artificial SBE promoters. Nevertheless, it has been shown that Smad2 function can also be antagonised by c-Jun. Transfection with c-Jun inhibits TGF-β-dependent activation of reporters driven by the ARE element, which is Smad2-dependent (Pessah et al., 2002; Pessah et al., 2001). Consistent with these findings, it has been shown that overexpressed c-Jun can interact with overexpressed c-Ski (Pessah et al., 2002) and TGIF (Pessah et al., 2001), both of which are Smad corepressors (see below). This provides a possible mechanism for inhibition of Smad function by c-Jun at specific promoters.

In all reports published thus far it seems that AP-1 factors can cooperate with Smads when an AP1 site is present in the promoter (with or without an SBE) (Kang et al., 2003; Qing et al., 2000; Wong et al., 1999; Zhang et al., 1998) and that they antagonise Smad function when an SBE is present without an AP-1 site (Dennler et al., 2000; Pessah et al., 2002; Pessah et al., 2001; Verrecchia et al., 2001a). There is no evidence that interaction with AP-1 family members can interfere with DNA binding by Smads. These findings require a model where c-Jun and JunB can interact with Smads independently of DNA and recruit specific corepressors to antagonise Smad function. When an AP-1 site is also present, these coprepressors must not be recruited and AP-1 factors or Smads must instead recruit coactivators that induce gene transcription.

As well as making indirect responses dependent on sustained TGF-β signalling, the cooperation of Smads with AP1 may allow integration of the JNK pathway and other signals with TGF-β. ATF3, for example, may integrate the TGF-β with stress signalling. Various stress signals and inflammatory cytokines such as TNF-α activate the JNK and p38 MAPK pathways and thus induce *ATF3*. Activation of stress signalling has been shown to cooperate with TGF-β in inducing *Id1* and both pathways cooperate in inducing cell-cycle arrest (Kang et al., 2003).

1.4.6 Smads and Sp1 – interaction with a ubiquitous transcription factor

A substantial volume of literature has dealt with the proposal that active Smad complexes cooperate with the ubiquitously expressed transcription factor Sp1. This was the first eukaryotic transcription factor to be identified and cloned (Dynan and Tjian, 1983). It binds DNA via three zinc fingers that recognise various GC-rich elements and regulates the expression of a wide range of genes by with different signalling pathways in different tissues (Safe and Abdelrahim, 2005). Other members of the Sp family include Sp2-Sp8 and the Kruppel-like factors (KLFs).

In overexpression studies, Sp1 has been found to physically associate with Smad2, Smad3 and Smad4 in a TGF-β-dependent fashion (Feng et al., 2000; Pardali et al., 2000) and at least three separate studies have confirmed that endogenous Sp1 associates with endogenous Smad complexes. However, the functional significance of this interaction remains poorly understood and has relied mainly on promoter analysis carried out by EMSA and reporter assay. Target genes that have been proposed as joint Smad–Sp1 targets include $p21^{Cip1/Waf1}$, $p15^{Ink4b}$, *Smad7*, *PAI-1*, *c-Met* and the genes encoding the Integrin β5 subunit, Tenascin C and α2(I) Collagen. Interestingly, TGF-β is overexpressed in Alzheimer's patients and active Smad3-Smad4 complexes have been reported to cooperate with Sp1 in the induction of β-amyloid precursor protein (β-APP) (Docagne et al., 2004). However neither chromatin immunoprecipiation nor siRNA knock down has been employed to confirm the importance of Sp1 in TGF-β-dependent control of any gene.

That Sp1 is ubiquitously expressed is apparently inconsistent with the model that Smad-interacting transcription factors are cell-type specific (Fig. 1.8A). However, the interaction may be modulated by post-translational modifications or by interaction with other factors that are expressed or activated in a more restricted fashion. In support of this latter view there is evidence that Sp1 and Smads cooperate with AP1 to induce *Smad7* (Brodin et al., 2000) and with Ets1 to induce *Tenascin-C* (Jinnin et al., 2004). It is also possible that Sp1 will be found to be involved mainly in the more "general" responses to TGF-β such as cell cycle arrest.

1.4.7 Interaction of Smads with transcription factors regulated by other pathways

As already discussed in the case of AP1, regulation of Smad transcriptional partners by extracellular signals can provide a mechanism to integrate TGF-β with other biochemical messengers. Similarly, PI3K signalling antagonises the function of Smad partners of the FoxO family (see section 1.4.8 below) (Seoane et al., 2004). Another example of this is the cooperation of Smads with Lef1/Tcf factors, which are controlled by Wnt–β-catenin signalling.

Members of the Lef1/Tcf family are transcription factors that are activated downstream of Wnt signalling and bind DNA via a high-mobility group (HMG) domain (Giese et al., 1991; Tolwinski and Wieschaus, 2004). Binding of Wnt ligands to Frizzled receptors indirectly activates the cytoplasmic protein Dishevelled. Active Dishevelled inhibits GSK3β, preventing phosphorylation of β-catenin and its recognition by a destruction complex containing GSK3β, Apc and Axin. β catenin then accumulates in the nucleus, binds Lef1/Tcf factors, displaces corepressors of the Groucho family and induces transcription (Tolwinski and Wieschaus, 2004).

Immunoprecipitation experiments using cultured mammalian cells suggest that Smad4 can associate with Lef1 and β catenin and that the interaction with Lef1 is independent of β catenin. However these experiments relied on overexpression and did not investigate the effect of TGF-β superfamily signals on the interactions (Labbe et al., 2000; Nishita et al., 2000). Functionally, Smads seem to cooperate in the

induction of the Wnt target genes *Siamois* and *Xtwn* and expression of these proteins is highest in the dorsal vegetal quagrant of the embryo, where both signals are active (Crease et al., 1998; Nishita et al., 2000). The promoter region of *Xtwn* contains four Smad-binding sites and, more proximally, four LEF/TCF sites. In reporter assays carried out in embryos, deletion of the Smad sites markedly reduces Wnt8-dependent transcription from the *Xtwn* promoter (Nishita et al., 2000). Conversely, Wnt signalling may collaborate in induction of the Nodal target genes *goosecoid* and *chordin* (Crease et al., 1998).

1.4.8 Smad-transcription factor interactions in tumourigenesis

Transcription factors that physically or functionally interact with Smads are likely to play a major role in tumourigenesis. In the case of tumour cells that have an intact TGF-β–Smad pathway, the simplest model to explain the selective loss of the cytostatic response is loss-of-function in transcription factors required for cytostatic gene responses but not for other responses. Loss of repression of the oncogenic transcription factor *c-Myc* is one event that fits into this category (Chen et al., 2001). Other transcription factors where there is reason to believe in such a model include FoxO family members, FoxG1 and p53 (Cordenonsi et al., 2003; Seoane et al., 2004). However, each of these requires further study to conclusively demonstrate a direct link to TGF-β in tumourigenesis.

Like FoxH1, other members of the forkhead superfamily – the FoxO family and FoxG1 – have been shown to be important in TGF-β signalling in cells derived from the mammalian adult (Seoane et al., 2004). The SIM and FM sequences of the Mix and FoxH1 families are not conserved in any of these factors (Germain et al., 2000) but, as with FoxH1, they have their DNA-binding domain at the N-terminus and a transactivation domain at the C-terminus (Kaestner et al., 2000; Kaufmann and Knochel, 1996). The $p21^{Cip1/Waf1}$ promoter contains a forkhead binding element adjacent to a Smad site and FoxO seems to cooperate with Smad3 and Smad4 in inducing this gene. A problem with this model is that, while very good interaction data was obtained with FoxO3, knock down of this factor did not impede $p21^{Cip1/Waf1}$

induction or cell-cycle arrest in response to TGF-β (Seoane et al., 2004). Multiple knock downs of members of the FoxO family will thus be required to confirm their functional significance.

The physical basis of the Smad-FoxO interaction appears to be very different from that of the Smad-FoxH1 interaction. It seems to be mediated by the DNA-binding domain of FoxO and the MH1 of Smad3 and Smad4 (Seoane et al., 2004). This suggests that the interactions of the Smads with FoxH1 and FoxO evolved separately.

FoxG1 seems to antagonise TGF-β induction of $p21^{Cip1/Waf1}$ by recruiting HDAC1 to the promoter. FoxG1 does not appear to interact directly with Smads or DNA to achieve this function but does it instead by binding to FoxO (Seoane et al., 2004). Other contexts where FoxG1 seems to antagonise Smad function is in the induction of *PAI-1* and $p15^{Ink4b}$ (Rodriguez et al., 2001). FoxG1 is expressed in the developing telencephalic neuroepithelium and in glioblastomas. In both settings it appears to make cells refractory to the cytostatic effect of TGF-β. This provides a fascinating insight into how deregulated expression of Smad transcriptional partners may be responsible for the altered response to TGF-β seen in many tumours. A further insight is provided by the facts that the PI3K pathway antagonises FoxO function and that this pathway is overactive in a wide range of cancers due to loss of PTEN phosphatase (Seoane et al., 2004).

The tumour suppressor and transcription factor p53 has also been shown to be involved in a range of TGF-β-induced gene responses. Knock down of p53 in HepG2 cells results in impaired Activin-dependent induction of $p21^{Cip1/Waf1}$, *PAI-1* and *MMP-2* but does not affect the induction of *TIEG* (Cordenonsi et al., 2003). Other genes that appear to be regulated by p53 in combination with Smads include *GADD45α* (Cordenonsi et al., 2003) and *AFP*. *AFP* encodes α-fetoprotein, an hepatic tumour marker that is cooperatively downregulated by TGF-β and p53 (Wilkinson et al., 2005). Functionally, the cytostatic effect of TGF-β is greatly impaired in $p53^{-/-}$ mouse embryonic fibroblasts (MEFs) compared to wild-type cells. p53 also seems to be required for Activin/Nodal signalling in *Xenopus* embryos. Morpholino knock-down

of p53 results in loss of expression of mesoendodermal markers induced by Activin/Nodal and disrupted gastrulation movements (Cordenonsi et al., 2003).

It seems that p53 can also antagonise Smad function. It appears to antagonise TGF-β-dependent induction of *COL1A2* and may inhibit matrix production in response to TGF-β (Ghosh et al., 2004). A solution-phase interaction of overexpressed Smad2 but not Smad4 with overexpressed p53 has been demonstrated. It has also been shown, by DNA pull-down, that Smad2, Smad3 and p53 all associate with the *GADD45α* promoter (Cordenonsi et al., 2003). It is thus likely that the functional cooperation of Smads with p53 reflects a physical association of these proteins.

p53 and TGF-β are both major tumour suppressors and all of this data suggests that a major theme in Smad-dependent transcriptional control may be interaction with factors that also function as tumour suppressors. Loss of such factors would provide a simple mechanism by which cytostatic gene responses to TGF-β could be compromised without affecting other responses. Other tumour suppressors that have been reported to interact with Smads include BRCA2 (Preobrazhenska et al., 2002) and it is possible that loss of this tumour suppressor may affect the TGF-β response in mammary carcinoma cells.

Another transcription factor that is involved in both TGF-β signalling and cell cycle regulation is E2F. However, no link to tumourigenesis has been shown. The E2F family consists of multiple partly redundant factors (E2F1-8) that function as heterodimers with bHLH factors of the DP family, usually DP1. The HLH protein Rb is the canonical E2F corepressor but p107 and p130 can also fulfil this role. Rb has a higher affinity for E2F1-3 whereas p107 and p130 are selective for E2F4-5 (Dimova and Dyson, 2005).

A pre-formed complex of Smad3, E2F4, E2F5, DP1 and p107 has been shown to exist in the cytoplasm in the absence of TGF-β. Upon activation of TGF-β receptors, this complex accumulates in the nucleus and binds to juxtaposed SBE and E2F elements in the *c-Myc* promoter. This results in rapid repression of *c-Myc* and combined knock

down of E2F4 and E2F5 abrogates the ability of TGF-β to repress this gene (Chen et al., 2002b).

These results are intriguing because E2F was previously known as the ultimate recipient of CDK signals in G_1 and as the core positive regulator of cell cycle progression. Induction of *c-Jun*, *c-Fos* and *c-Myc* by mitogens results in delayed induction of D-type cyclins, Cdk2, Cdk4 and E2Fs. Cyclin D-Cdk4/6 complexes then phosphorylate and inhibit Rb with the result that E2Fs are able to auto-induce their own transcription along with activating *Cyclin E*, *Cdk2* and genes such as *TK* and *DHFR* that play a mechanistic role in S-phase (Chan et al., 2001). The finding that E2F4 and E2F5 cooperate with TGF-β to repress *c-Myc* shows that they play a more complex role in cell cycle control, acting in a positive fashion downstream of CDK and in a negative fashion upstream of it (Chen et al., 2002b). E2F may have evolved these opposing functions to suppress the oncogenic effect of E2F4/E2F5 overexpression. In a similar fashion, E2F overexpression has been shown to induce apoptosis (Paulson et al., 2006), a function that may serve to prevent E2F from leading to tumourigenesis.

1.4.9 Complex regulation of TGF-β dependent promoters

Smads have been reported to interact with multiple transcription factors on the promoters of TGF-β target genes. One example of this is the $p21^{Cip1/Waf1}$ promoter, where they have been reported to interact with Sp1, FoxO and p53. Another is *PAI-1*, where Sp1 and TFE3 may be important Smad partners. In each case, the relevant Smad-responsive sites are distinct and, though their functional relationship has not been investigated, it is possible that they interact to allow a precise level of transcription. It is also possible that the different TGF-β responsive elements present in a single promoter are active in distinct physiological states, reflecting the availability of the relevant transcription factor partners. These are issues that require further investigation and which may lead to further insights into Smad-transcription factor cooperation.

1.4.10 The requirement for Smad4 in TGF-β gene responses

All of the Smad-transcription factor interactions discussed thus far seem to involve interaction of various partners with active Smad complexes that contain Smad4 along with R-Smads. However, there may exist functional Smad-transcription factor complexes lacking Smad4. This possibility is supported by the fact that R-Smad dimers and trimers can form in the absence of Smad4 and undergo nuclear accumulation (Fink et al., 2003) and by the finding that TGF-β can induce a range of target genes in Smad4-null cell lines (Subramanian et al., 2004). Similarly, a definitive microarray analysis carried out using cell lines inducibly expressing an shRNA against Smad4, revealed the existence of a large number of Smad4-independent TGF-β target genes (Levy and Hill, 2005). Many of these genes are rapidly induced or repressed (Levy and Hill, 2005) and it is possible that they are controlled by Smad-independent pathways activated directly from the TGF-β receptor. However, some of them may instead be activated by R-Smads independently of Smad4. This possibility is supported by the more severe phenotype of *mad* (R-Smad homolog) mutants compared to *medea* (Smad4 homolog) mutants in *Drosophila* (Wisotzkey et al., 1998). It is also consistent with the finding that conditional inactivation of *Smad4* in the early mouse embryo abrogates some developmental events that are dependent on the TGF-β superfamily but does not affect others (Chu et al., 2004).

1.4.11 Smad interference with the function of other transcription factors

An entirely different mechanism by which Smads may operate is direct interference with the function of other transcription factors. Two proposed instances of this relate to muscle differentiation, a process that is repressed by TGF-β. Activated Smad3 has been reported to bind the HLH domain of MyoD – the classical muscle specification factor – and to prevent its interation with E12 and E47, thus blocking the myogenic programme (Liu et al., 2001). Similarly, activated Smad3 may interact with MEF2 and block its interaction with the coactivator GRIP1 (Liu et al., 2004).

1.5 Smad–coregulator interactions

Once they are bound to target gene promoters, Smad-transcription factor complexes must initiate transcription activation or repression. Smads achieve this function by recruiting transcriptional coregulators – either independently or in cooperation with their associated transcription factors (Massague et al., 2005). The transcriptional coactivators recruited by Smads function by recruiting factors that open the chromatin template and allow target gene transcription (Ross, S.J. *et al*, manuscript submitted) (Massague et al., 2005). Smad corepressors operate in the converse way, by recruiting factors to close the template. These corepressors function both in gene repression by activated Smad complexes and to prevent "inadvertent" gene activation by inactive Smads (Massague et al., 2005). The remainder of this section will review the mechanisms of Smad-mediated transcriptional control that are have been well characterised.

1.5.1 Smad-mediated activation of transcription

In vitro studies indicate that Smad complexes recruited to DNA by the Gal4 DNA-binding domain fused to an FM-SIM sequence cannot activate transcription on a naked DNA template. This indicates that they are incapable of independently recruiting the basal transcription apparatus. On chromatin, however, Smad complexes recruited in this way can activate transcription and appear to do so by recruiting CBP/p300, Brg1 and other coregulators involved in histone modification or chromatin remodelling (Ross, S.J. *et al*, manuscript submitted) (Shi and Massague, 2003). All of this suggests a model in which the main function of Smads in gene activation is to open the chromatin template for transcription. Once this has been achieved the basal transcription apparatus may be recruited either by Smad-associated transcription factors or by factors that bind histone tails modified downstream of the Smads. A candidate for this role would be $TAF_{II}250$, which can bind acetylated lysines through its double bromodomain (Jacobson et al., 2000).

The Smad coactivators that have received most attention are the histone acetyltransferases CBP and p300. These have been reported to interact with the MH2 domains of Smad1, Smad2, Smad3 and Smad4 and to be involved in the induction of *PAI-1*, *COL1A2*, $p15^{Ink4b}$ and $p21^{Cip1/Waf1}$ (Feng et al., 1998; Pouponnot et al., 1998; Shen et al., 1998; Topper et al., 1998). p300 has also been reported to be recruited to Smad4 by the coactivator SMIF, which associates with the Smad activation domain (SAD) in the linker region of Smad4 and undergoes nuclear accumulation in parallel with Smad4 (Bai et al., 2002). Another Smad coactivator that also interacts with both the SAD and p300 is MSG1 (Shioda et al., 1998; Yahata et al., 2000). The importance of p300 in TGF-β-dependent gene activation is clearly demonstrated by *in vitro* assays in which addition of the p300 inhibitor Lys-CoA completely abolishes Smad-mediated transcription on a chromatin template (Ross, S.J. *et al*, manuscript submitted).

Interestingly, the adenoviral oncoprotein E1A inhibits Smad TGF-β function and may do so by sequestering p300 – providing another mechanism by which tumour cells may become insensitive to the cytostatic effects of TGF-β (Nishihara et al., 1999). p300 may also provide a platform for the integration of TGF-β with other signals. The cooperation of BMP4 and LIF (leukaemia inhibitory factor) in inducing *GFAP*, for example, is mediated by the ability of Smad1 and Stat3 to collaborate in recruiting p300.

Other coregulators reported to interact with Smads include Brg1, PCAF, Gcn5 and the mediator subunit ARC105 (Massague et al., 2005). It is perhaps significant that PCAF is commonly associated with p300 (Yang et al., 1996) since this suggests that they may be recruited simultaneously to TGF-β-responsive promoters. The interaction of Smad complexes with ARC105 (Kato et al., 2002) is in apparent discord with the finding that Smads cannot independently recruit the basal transcription apparatus (Ross, S.J. *et al*, manuscript submitted). However, it remains possible that Smads do interact with the basal apparatus via mediator" but remain dependent on other factors for efficient recruitment.

1.5.2 Smad-mediated repression of transcription

At least a quarter of all direct TGF-β gene responses are repression events that result from Smad-mediated recruitment of corepressors (Massague et al., 2005). One such mechanism of Smad-dependent gene repression is direct and indirect recruitment of histone deacetlyases – mirroring what occurs with p300. It has already been mentioned that p107 is recruited by Smads to the *c-Myc* promoter and that this results in *c-Myc* repression. p107 achieves this function by recruiting HDAC1 (Chen et al., 2002b). Similarly, Smad3 has been reported to directly recruit HDAC4 to repress Runx2-mediated induction of *Osteocalcin* in osteoblasts (Kang et al., 2005). Runx2 is a transcription factor that interacts with Smad3 and which can also interact directly with HDAC4 – suggesting that Smad3 and Runx2 may create a high affinity/avidity binding site for HDAC4 (Vega et al., 2004) .

Other Smad corepressors seem to work as general antagonists of Smad function. Two examples are TGF-β interacting factor 1 (TGIF1) and TGIF2. These interact with Smad2 and recruit the Sin3/HDAC complex via the Sin3 and the adaptor CtBP. TGIF has also been shown to interfere with recruitment of p300 to active Smad complexes. Antisense interference with TGIF1 results in elevated expression of PAI-1 and overexpression interferes with TGF-β-dependent activation of an ARE reporter, suggesting that this corepressor has a general role in regulating the level of Smad activity (Melhuish et al., 2001; Wotton et al., 1999a; Wotton et al., 1999b). TGF-β signalling may also induce *TGIF1*, providing the possible basis for another negative feedback loop regulating the pathway. In addition, TGIF is phosphorylated by ERK at two distinct sites, leading to stabilisation of the protein and antagonism of TGF-β function (Lo et al., 2001).

Along with this general role it is possible that TGIF functions as a promoter-specific Smad corepressor. One reported example of this is downregulation of *AhR* (aryl hydrocarbon receptor) in human lung cancer cells. Functionally, a human mutation preventing interaction of TGIF with CtBP causes severe defects in craniofacial development (Melhuish and Wotton, 2000). This is reminiscent of the phenotype of

TGF-β3-null mice (Kaartinen et al., 1995; Proetzel et al., 1995) and suggests that a precise level of TGF-β signalling is required for proper craniofacial patterning.

Perhaps the best-characterised Smad corepressors are members of the c-Ski family of proto-oncoproteins. *v-Ski* was first identified as the transforming agent of the avian Sloan-Kettering retrovirus which induces oncogenesis in chick embryonic cells. The human cellular homolog, *c-Ski*, and a close family member, *SnoN*, were subsequently cloned and shown to also be oncogenic. However, it was only later that these proteins were found to function by inhibiting Smad-mediated gene activation (Luo, 2004).

c-Ski and SnoN bind to defined regions of the MH2 domain of Smad2 and Smad3 and a distinct region on the surface of the MH2 domain of Smad4 (Akiyoshi et al., 1999; Luo et al., 1999; Stroschein et al., 1999; Sun et al., 1999). In addition, c-Ski but not SnoN can interact with Smad1 and Smad5 (Wang et al., 2000). Importantly, the regions of Smad2/3 and Smad4 bound by these proteins overlap the regions that form the R-Smad–Co-Smad interface. A crystal structure of the Smad4 MH2 domain in complex with a c-Ski fragment is available. In this structure the Smad-binding fragment of c-Ski adopts a SAND domain conformation and chelates two zinc ions. A specific loop – the I-loop – on the surface of c-Ski makes specific contacts with the L3 loop of Smad4 (Wu et al., 2002).

c-Ski and SnoN repress Smad function in two ways. Firstly, by binding at the Smad-Smad interface they disrupt the formation of activating complexes and may instead mediate the formation of alternative repressive complexes (Luo, 2004). Secondly, these factors have been shown to recruit the Sin3-HDAC complex by interacting with Sin3 and the adaptor N-CoR (Luo et al., 1999; Stroschein et al., 1999).

Functionally, c-Ski and SnoN seem to act as general repressors of Smad function in a similar fashion to TGIF. They are expressed at low levels in the absence of TGF-β and are rapidly degraded upon addition of ligand. In the case of SnoN at least this occurs via the Smad3-mediated recruitment of the anaphase promoting complex and the ubiqitin ligase Smurf2. These poly-ubiqitinate SnoN, leading to its degradation by the proteasome (Bonni et al., 2001; Stroschein et al., 1999; Sun et al., 1999).

However, TGF-β also induces SnoN at the transcriptional level and the proteins begin to accumulate after 1.5-2 hours of signalling (Stroschein et al., 1999). All of this suggests a model in which – like FKBP12 and the need to process TGF-β ligand – the c-Ski family function to prevent "inadvertent" pathway activation in the absence of signal. TGF-β-dependent degradation and re-synthesis then allow a time window during which maximal TGF-β gene responses are possible followed by a longer phase of more limited signalling. As with TGIF, c-Ski and SnoN may also be found to have promoter-specific functions in TGF-β signalling.

A final Smad coprepressor that has been well characterised is Smad nuclear inhibitory protein 1 (SNIP1). This has been reported to interact with Smad1, Smad2 and Smad4 and to prevent recruitment of CBP/p300. As with c-Ski, SnoN and TGIF, this protein appears to be another general antagonist of Smad function (Kim et al., 2000).

1.6 Smad–transcription factor interactions on the *c-Jun* promoter

Despite the fact that a substantial number of Smad transcriptional partners have been reported, it is likely that many remain to be discovered. One reason for reaching this conclusion is that many of the reported interactions have not been thoroughly validated. Another is that, given the functional pleiotropy of the Smads, a large number of partners are to be expected. In addition, most of the Smad corepressors that have been characterised function as general antagonists of Smad function rather than being recruited by Smads to selectively repress specific genes. It is likely that more specific corepressors await discovery. Finally, no SIM-containing factor has been identified in cells derived from mammalian adults but there is evidence that functionally equivalent factors are present in this context (see below) (Inman and Hill, 2002). A system that may allow identification of additional Smad transcriptional partners is the *c-Jun* promoter.

c-Jun is a direct target gene of TGF-β and undergoes a strong transient induction in mammalian cells (Pertovaara et al., 1989). The *c-Jun* promoter region has been used as a model to study transcriptional control by TGF-β. Previous work has shown that a dimer of Smad3 and Smad4 associates with a Smad Binding Region (SBR) located in the 5'-UTR of

this gene and can thus activate its transcription. This 55 base-pair SBR is sufficient for TGF-β inducibility in luciferase reporter assays but an AGAC triplet located within this region and known as the Smad Binding Element (SBE) has been identified as the actual binding site of the Smads (Fig. 1.9A)(Inman and Hill, 2002; Wong et al., 1999).

It has also been shown that at least two additional factors associate with Smad3 and Smad4 on the c-Jun SBR. These factors remain unidentified but it is known that one is a DNA-binding protein that associates with a CCAG triplet downstream of the SBE and the other is a SIM-containing factor that is likely not to interact directly with DNA. The SIM-containing component has been temporarily designated factor X and the CCAG-binding protein factor Y (Fig. 1.9B) (Inman and Hill, 2002).

The existence of these factors was determined by electrophoretic mobility shift assay (EMSA) carried out using radiolabelled probe corresponding to the wild-type *c-Jun* SBR or to an SBR with the CCAG triplet mutagenised. Complex formation is reduced on the mutant probe compared to wild-type, a result that is also obtained with a truncated probe in which the CCAG triplet and all downstream base-pairs are deleted. This CCAG triplet does not function as a degenerate Smad binding site since its addition in large excess as a nonlabelled competitor in EMSA does not affect complex formation, whereas addition of an unlabelled SBE disrupts the complex. These data indicate that a Smad partner that binds specifically to a CCAG triplet is present in this complex and that it cannot interact with the probe without the cooperation of the Smad3-Smad4 dimer. This factor has been termed factor Y (Fig. 1.9B) (Inman and Hill, 2002).

In the same assays, addition of a peptide corresponding to the Mixer SIM destabilises the complex and causes a slight increase in electrophoretic mobility, an effect that is not seen when critical residues in the peptide are mutagenised. The SIM peptide does not to disrupt the Smad3-Smad4 interaction, leading to the conclusion that a second unidentified factor is present in the complex and that it interacts with Smad3 because it contains a sequence functionally equivalent to the Mixer SIM. This second factor has been termed factor X (Fig. 1.9B). Factor X is thought not to associate directly with DNA and is therefore likely to be a transcriptional coregulator (Inman and Hill, 2002).

Use of a mutagenised or truncated probe together with addition of SIM peptide have an additive destabilising effect on the complex, indicating that X and Y are likely to be distinct factors and not the same protein. Binding sites similar to that for factor Y occur in the promoter regions of several target genes of the TGF-β superfamily, including *Smad7* and *PAI-1*, and addition of SIM peptide disrupts complex formation on EMSA probes corresponding to *Smad7* SBR. These observations indicate that factors X and Y, or related factors, are likely to be of general importance in TGF-β signalling (Inman and Hill, 2002).

It has not been possible to predict the identity of factor X on the basis of sequence similarity to known SIM-containing factors due to our limited understanding of which amino acid sequences can constitute a functional SIM. Similarly, testing a candidate for factor Y; nuclear factor 1 (NF1), for which a weak consensus binding site overlaps the CCAG triplet has given negative results. Instead, my project is aimed at identifying these factors by affinity purification and mass spectrometry.

CHAPTER 2

MATERIALS AND METHODS

2.1 Introduction

The following sections describe the methods employed in chapters three and four of this thesis. The materials required for each method are described in the corresponding sections. All materials were of analytical grade and, unless stated otherwise, were purchased from Sigma-Aldrich, BDH laboratory supplies or Fisher Scientific. All water used was deionised on a Milli-Q plus system (Millipore) and, where appropriate, buffers were sterilised by filtration on a 0.2 μm vaccum-driven filtration system (Stericup).

2.2 Cell culture

HaCaT cells (Cancer Research UK Cell Services) were cultured on plastic dishes, wulti-well plates or flasks of tissue culture grade (Corning Incorporated and Beckton Dickinson) in high-glucose DMEM (Cancer Research UK Cell Services) supplemented with 10% foetal calf serum (Gibco or Invitrogen) and warmed to 37 °C before addition to cells. Cultivation was carried out in an incubator (Sanyo) at 37 °C and 10% CO_2 and cells were subcultured when they reached 80-90% confluence. To remove cells from the flask, they were washed with PBSA (Cancer Research UK Cell Services) and incubated for 10 minutes with trypsin-versene solution (Cancer Research UK Cell Services). Subculturing was carried out by dilution of the trypsinised cells in culture medium and aliquoting into a fresh flask or dish.

2.3 Generation of stable cell lines

HaCaT cells were grown to approximately 50 % confluence on a 6-well tissue culture plate (Corning) before transfection by the FuGENE 6 procedure (Roche). For each

well, 3 μl of FuGENE reagent were combined with 90 μl Opti-MEM I reduced serum medium (Invitrogen) and allowed to stand for 5 minutes in a tissue culture cabinet. 0.1 μg pSuperRetro (a gift from Julian Downward) and 0.9 μg pUC12-EF-HA-Smad3 (generated by Gareth J. Inman) were then added and the mixture was allowed to stand for a further 15 minutes. Culture medium was then removed from the cells and 1 ml fresh medium was added. The transfection mixture described was then added to the cells and the cells were placed in an incubator at 37 °C and 10% CO_2 overnight. Cells were then subcultured onto 15 cm dishes and allowed to grow to 80-90 % confluence. HaCaT cells transfected and subcultured in this way were then selected for puromycin resistance. The culture medium was removed and replaced with medium supplemented with 1 μg ml^{-1} puromycin. Cells were then placed in an incubator at 37 °C and 10% CO_2 and grown until non-resistant cells had died and resistant colonies had emerged. This required replacement of puromycin-supplemented medium at three-day intervals. Cloning cylinders were prepared by cutting a 9-10 mm section from the broad end of a 1 ml pipette tip and covering the uncut end with vacuum wax. These were placed, waxed end downwards over individual colonies to form a seal around the colonies. A drop of trypsin-versene was then placed into each cylinder and the dish was placed in an incubator for 10 minutes. The liquid was then drained from each cylinder and placed into one well of a 24-well tissue culture plate. Fresh medium supplemented with 1 μg ml^{-1} puromycin was then added and the cells were placed in an incubator at 37 °C and 10% CO_2. Cells surviving this process were further subcultured into 6-well plates.

2.4 TGF-β inductions

TGF-β inductions were carried out by adding TGF-β to the medium of cultures grown to 80-90 % confluence. Stock TGF-β (PreproTech) was diluted to 1 μg ml^{-1} in 4 mM HCl and 1 mg ml^{-1} bovine serum albumin and was added to cultures at 1 in 500 dilution to give a final concentration of 2 ng ml^{-1} TGF-β. Cells were then placed in an incubator at 37 °C and 10% CO_2 before harvesting of cells for nucleocytoplasmic fractionation or cell fixation for immunocytochemistry.

2.5 Immunocytochemistry

Immunostaining of cultured cells was carried out by the method of Pierreux *et. al.* (Pierreux et al., 2000). Microscope coverslips were placed into the wells of 12- or 24-well tissue culture plates and cells were subcultured into these wells. Cells were then either induced with TGF-β or left uninduced. Culture medium and TGF-β were then washed away with three 1-3 ml volumes of PBSA and cells were fixed for 15 minutes at room temperature with 3.7 % formaldehyde in PBSA. Cells were then washed three times with 1-3 ml PBSA and permeabilised for 10 minutes with 0.3 % triton X-100 in PBSA. Blocking was subsequently carried out for 30 minutes at room temperature with a solution containing 0.3% avid bovine serum albumin, 10 % foetal calf serum (Invitrogen), 5 % powdered milk (Marvel) and 0.3 % triton X-100 in PBSA. Mouse monoclonal anti-HA antibody 12CA5 was then added at 5 $\mu g\ ml^{-1}$ in blocking solution without milk and incubated for 1 hour at room temperature without agitation. Cells were then washed three times with 1-3 ml of 0.1 % triton X-100 in PBSA. Goat polyclonal anti-mouse IgG conjugated to Texas red was added at a dilution of 1 in 200 in blocking solution without milk and cells were incubated for 30 minutes at room temperature without agitation. Antibody was washed away with three volumes of 0.1 % triton X-100 in PBSA. Cells were then fixed for 15 minutes at room temperature with 3.7 % formaldehyde in PBSA, washed three times with 0.1 % triton X-100 in PBSA and finally washed three times in water. Single drops of Mowiol mounting solution were placed at the centre of microscope slides and a coverslip bearing stained and fixed cells was inverted over each drop so that the cells lie between the slide and the coverslip.

2.6 Microscopy

Immunostained cells were visualised using a Zeiss 520 laser scanning confocal light microscope. Samples were excited at 543 nm and observed with a long-pass filter allowing passage of wavelengths of 600 nm and above.

2.7 Extract of cells

Cells were grown to 90% confluence in 10 cm tissue culture dishes and extract was prepared by the method of Marais *et. al.* (Marais et al., 1993). Cells were washed twice with ice cold PBSA and scraped from tissue culture dishes into 250 µl D 0.4 buffer. This buffer consisted of 50mM HEPES pH 7.5, 10% (v/v) Glycerol, 0.4M KCl, 0.4% triton, 10mM EGTA, 5mM EDTA, 1 mM dithiothreitol, 10 µg ml^{-1} leupeptin, 20 µg ml^{-1} bestatin, 10 µg ml^{-1} E-64 protease inhibitor, 20 µg ml^{-1} pepstatin, 10 µg ml^{-1} aprotinin, 0.4 mM pefabloc SC and 0.5 mM benzamidine. Cells were then sonicated for 10 seconds on ice and extracts were centrifuged for 10 minutes at 12,500 rpm and 4 °C in a Biofuge Fresco Heraeus centrifuge (Sorvall). Extracts were stored at -80 °C until use.

2.8 Nuclear extract of cells

Nuclear proteins were extracted essentially by the method of Wong *et. al.* (Wong et al., 1999). Cells were grown to 90% confluence in tissue culture dishes. Medium was removed and cells were washed twice with ice-cold PBSA, ensuring that all PBSA was removed after the second wash. Hypotonic buffer was then added to disrupt the plasma membrane by osmotic shock, 1ml of hypotonic buffer to 15 cm dishes and 0.5 ml to 10 cm dishes. Hypotonic buffer consisted of 20 mM HEPES pH 7.6, 20 % (v/v) glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1 % (v/v) triton X-100, 25 mM NaF, 25 mM sodium β-glycerophosphate, 1 mM dithiothreitol and one complete protease inhibitor tablet (Roche) per 40 ml of buffer. Dithiothreitol and protease inhibitors were added fresh to this buffer. Cells were then scraped off the dish and centrifuged for 15 minutes at 1000 rpm and 4 °C in a Beckman Allegra 6KR swinging-bucket centrifuge. The supernatant corresponding to the cytoplasmic extract was then disposed of and the pellet consisting of intact nuclei was washed by resuspension in hypotonic buffer and re-pelleted by centrifugation for 5 minutes at 1000 rpm and 4 °C. The pellet was then resuspended in hypertonic buffer (same composition as hypotonic but with 500 mM NaCl) and allowed to lyse for 1-3 hours at 4 °C with gentle agitation. Nuclear lysates were then centrifuged for 20 minutes at

12,500 rpm and 4 °C in a Biofuge Fresco Heraeus centrifuge. Supernatants were removed, protein concentrations were determined and the extract was stored at -80 °C.

2.9 Determination of protein concentrations

Protein concentrations were determined by the Bradford method. A standard curve was created by combining 1, 2, 3, 5 and 8 µl aliquots of 1 µg μl^{-1} bovine serum albumin with 200 µl of a 1 in 5 dilution of BioRad Protein Assay solution and measuring the absorbance at 595 nm on a SpectraMax Plus spectrophotometer (Molecular Devices). 1 µl extract samples were then combined with 200 µl of a 1 in 5 dilution of BioRad Protein Assay solution, the absorbance was measured at 595 nm and the protein concentration was determined from the standard curve.

2.10 DNA pull-downs

The following are the conditions under which analytical DNA pull-downs (DNAP) were carried out prior to optimisation for preparative purposes. They are based on the method of Kang *et. al.* (Kang et al., 2003). Double-stranded oligonucleotide baits were prepared by annealing 0.5 µg μl^{-1} of each of the appropriate single-stranded oligonucleotides in 1X buffer 2 from New England Biolobs. These annealing reactions were heated at 95 °C for 10 mins and allowed to cool slowly to room temperature over a period of 2 hours. Oligonucleotide sequences used were GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGG TCGGCA (*c-Jun* SBR wild-type) and GGAGGTGCGCGGAGTCAGGCATATATA TATATACAGCATGCATGCATGGTCGGCA (*c-Jun* SBR mutant). Both were 5' biotinylated on the strands shown. Non-biotinlyated competitor oligonucleotide with the same sequence as the mutant bait was also prepared. Neutravidin-coeated beads (Perbio) were then washed three times in "hypotonic" buffer with 140 mM NaCl instead of 10 mM NaCl and resuspended in a 10-fold excess of this buffer. Annealed oligonucleotide bait was then added (5 µg per 15 µl beads) and allowed to bind to the

beads for 1 hour at 4 °C with gentle agitation. Beads were then washed three times with 140 mM buffer and resuspended as a 1:1 slurry in this buffer. HaCaT nuclear extracts were adjusted to a protein concentration of 1 μg μl^{-1} by dilution in hypertonic buffer (starting protein concentrations were normally 1.1 to 1.3 μg μl^{-1}). 200 μl of extract were then diluted again by combination with 500 μl of hypotonic buffer to give a final NaCl concentration of approximately 140 mM. Additions were then made in the following order to DNAP reactions: 20 μg non-biotinylated mutant competitor oligonucleotide, 50 μl oligonucleotide-coated bead slurry and 700 μl diluted extract. DNAPs were then incubated overnight at 4 °C and washed three times in hypotonic buffer with 140 mM NaCl. Elution involved removal of the wash buffer, addition of 60 μl SDS sample buffer (contains 7.7 % [w/v] sodium dodecyl sulphate (SDS), 0.19 M Tris-HCl [pH 6.8], 37 % glycerol and 0.38 % β-mercaptoethanol) and heating at 95 °C for 5 minutes. Samples were then analysed by SDS-PAGE as described in 2.14 below, followed by immunoblotting as described in 2.15 below or by silver staining as described in 2.16 below.

2.11 Immunoprecipitations

The following are the conditions under which analytical immunoprecipitation (IP) were carried out prior to optimisation for preparative purposes. They are based on the method of Pierreux *et. al.* (Pierreux et al., 2000). Protein G- or protein A-sepharose beads (Sigma-Aldrich or Cancer Research UK Antibody Production) were washed and resuspended in a 10-fold excess of hypotonic buffer containing 140 mM NaCl. Appropriate antibodies were then added at a ratio of 5 μl antibody solution to 15 μl bead volume and beads were subjected to gentle agitation for 1 hour at room temperature. Antibodies used were 12CA5 mouse monoclonal anti-HA (Cancer Research UK Antibody Production) and rabbit polyclonal anti-Smad3 (Zymed or a gift from Peter ten Dijke). Beads were then washed three times and resuspended as a 1:1 slurry in hypotonic buffer containing 140 mM NaCl. Extracts were diluted as in 2.10 above. The following order of additions to IP reactions was then made: 5 μg wild-type or mutant oligonucleotide bait (where appropriate, sequences as in 2.10 above), 20 μg mutant competitor oligonucleotide (where appropriate, sequences as in

2.10 above), 30 µl bead slurry and 700 µl diluted extract. IPs were then incubated overnight at 4 °C with gentle agitation and washed three times in hypotonic buffer with 140 mM NaCl. Elution involved removal of the wash buffer, addition of 60 µl SDS sample buffer as described in 2.10 above and heating at 95 °C for 5 minutes. Samples were then analysed by SDS-PAGE and immunoblotting as described in 2.14 and 2.15 below.

2.12 Covalent crosslinking of anti-HA antibody to protein G-sepharose beads

For two-step affinity purification, the mouse monoclonal anti-HA antibody 12CA5 was covalently conjugated to protein G beads using dimethlypimelimidate (DMPD). The method used is based on that described by Harlow and Lane (Harlow, 1988). 500 µl dry beads were washed three times and resuspended in an excess of hypotonic buffer containing 140 mM NaCl. Antibody was then added at a ratio of 0.07 µg antibody per µl bead slurry and this mixture was subjected to gentle agitation at room temperature for 1 hour. Beads were then poured into a Bio-Spin disposable chromatography column (BioRad), washed with 5-7 ml of 0.1 M sodium borate (pH 9.0) and resuspended in 0.1 M sodium borate with 20 mM DMPD. This mixture was subjected to gentle agitation for 30 minutes at room temperature and poured into another disposable column. The beads were then washed with 5-7 ml of 0.2 M ethanolamine, resuspended in this buffer and subjected to gentle agitation for 2 hours at 4 °C. This mixture was then poured into another column, washed with 5-7 ml PBSA and stored in PBSA with 0.1 % sodium azide.

2.13 Two-step affinity purification

The following describes fully optimised two-step purification with 2 mg total protein as starting material. For the 30 mg scale, all quantities were scaled up by a factor of 15. It was important that all operations were carried out at 4 °C. Nuclear extract was diluted in the same fashion described in 2.10 but on a 10-fold larger scale, yielding 2 mg total protein in a volume of 7 ml hypotonic buffer containing 140 mM NaCl.

Beads prepared as described in 2.12 were washed three times and resuspended as a 1:1 slurry in hypotonic buffer containing 140 mM NaCl. Neutravidin-agarose beads were also washed three times and resuspended as a 1:1 slurry in hypotonic buffer with 140 mM NaCl. The following order of additions to step 1 reactions was then made: 5 μg biotinlyated oligonucleotide bait (wild-type, sequence as in 2.10 above), 25 μg non-biotinlyated competitor oligonucleotide (sequence as for mutant in 2.10 above), 60 μl HA bead slurry and 7 ml diluted extract. Step 1 reactions were then incubated for 4 hours with gentle agitation and poured into a Bio-Spin disposable chromatography column (BioRad). The beads were then washed with 1 ml hypotonic buffer containing 140 mM NaCl (33 bed volumes) and eluted by 5 sequential 8-minute incubations with 200 μl of 0.3 μg μl^{-1} HA peptide (Cancer Research UK Peptide Synthesis) in hypotonic buffer containing 140 mM NaCl. The sequence of this peptide was YPYDVPDYA. In step 2 of the purification, this eluate was added to 60 μl of neutravidin beads that had been washed three times in hypotonic buffer containing 140 mM NaCl. This mixture was then incubated overnight with gentle agitation. The mixture was then poured into a disposable column and washed with 1 ml hypotonic buffer containing 140 mM NaCl. To elute step 2, washed beads were incubated for 1 hour with 100 μl of hypotonic buffer containing 600 mM NaCl. Samples were taken for analysis by immunoblot or EMSA at various stages of this procedure as indicated in Figs. 4.8 and 4.9. In Fig. 4.8, samples of input and step 1 flow through for analysis by immunoblotting had a volume of 72 μl. Concentration to 30 μl on a Microcon centrifugal ultrafiltration system (Millipore) was thus required before these samples could be run on SDS-PAGE.

2.14 SDS-PAGE

Denaturing polyacrylamide gel electrophoresis with sodium dodecyl sulphate as the denaturing agent (SDS-PAGE) was carried out by the method of Laemmli (Laemmli, 1970). Two glass plates were assembled with silicon tubing to form a seal and 0.8 mm plastic spacers. 15 ml resolving gel mix was prepared consisting of 15 % (w/v) acrylamide, 0.0086 % (w/v) bisacrylamide and 0.375 M Tris-HCl (pH 8.8). 45 μl of each of TEMED and 20 % (w/v) ammonium persulphate were added to this mix. The

mix was then poured into the space between the glass plates to a height of approximately 70 % that of the plates and overlaid with 70 % ethanol. 5 ml stacking gel mix was then prepared consisting of 5 % (w/v) acrylamide, 0.0132 % (w/v) bisacrylamide and 0.125 M Tris-HCl (pH 6.8). 15 µl of each of TEMED and 20 % (w/v) ammonium persulphate were added to this mix. Ethanol was then removed from the top of the resolving gel, the gel was washed with water, the stacking gel was overlaid on top of the resolving gel and a comb consisting of an appropriate number of teeth was inserted between the glass plates. A further 15 minutes were then allowed for polymerisation. Once the gel had polymerised, the comb was removed and the gel and plates were clamped into an electrophoresis apparatus (Cambridge Electrophoresis Ltd.). Running buffer consisting of 0.38 M glycine, 0.05 M Tris base and 1 % sodium dodecyl sulphate was added and wells were thoroughly washed to free them of debris. Samples consisting of DNAP or IP eluates (as described in 2.10 and 2.11 above) or of protein samples combined with one-third their volume of SDS sample buffer (as described in 2.10 above) were heated for 5 minutes at 95 °C and placed in the wells. Molecular weight marker mix (produced by Mike Howell) was loaded in a well to one side of the gel. This contained 50 µg ml^{-1} each of myosin (212 kD), β-galactosidase (116 kD), phosphorylase (97.5 kD), bovine serum albumin (66.5 kD), catalase (57.8 kD), glutamate dehydrogenase (55.5 kD), ovalbumin (42.7 kD), carbonic anhydrase (28.9 kD), SBTI (20 kD) and RNaseA (13 kD) in SDS sample buffer diluted 1 in 4. Gels were then run for 1 hour and 30 minutes at 230 V and 65 mA prior to immunoblotting or silver staining of resolved proteins.

2.15 Immunoblotting of SDS-PAGE gels

Electrophoretic transfer was carried out by semi-dry method based on that described bt Harlow and Lane (Harlow, 1988). After being run, SDS-PAGE gels were soaked in semi-dry transfer buffer for 10 minutes. This buffer consisted of 0.15 M glycine, 20 mM Tris-HCl, 0.1 % (w/v) SDS and 20 % methanol. Six pieces of Whatman 3MM paper and a piece of Immobilon-P PVDF 0.45 µm membrane (Millipore) were cut to the size of the gel. The membrane was immersed in methanol for 30 seconds and, along with the whatman paper, immersed in transfer buffer for 1-2 minutes. A

sandwich was assembled in a SemiPhor transfer unit (Hoefer) consisting of three pieces of Whatman paper on the bottom, followed by the membrane, followed by the gel and followed by three additional pieces of Whatman paper on top. These were rolled over with a 10 ml pipette tube to ensure removal of all air bubbles. The lid of the apparatus was then put in place and proteins were transferred for 1 hour at 0.8 mA cm^{-2} of gel surface. After transfer the membrane was washed in PBST (PBSA with 0.2 % Tween) and stained with Ponceau S solution (2 % [w/v] Ponceau S and 30 % [v/v] trichloroacetic acid) to visualise the markers, allowing their position to be marked with a pen one excess stain has been washed away with water. Membranes were then blocked for 15 minutes with a 5 % powdered milk (Marvel) solution in PBST. Primary antibody was diluted in the same solution at an appropriate concentration and membranes were incubated with antibody overnight at 4 °C with gentle rocking. Primary antibodies used and the appropriate dilutions for each are as follows: mouse monoclonal anti-Smad2/Smad3 (cat. no. 610843) from BD biosciences (1 in 2000), B8 mouse monoclonal anti-Smad4 from Santa Cruz Biotechnology (1 in 2000), rabbit polyclonal anti-Phospho-Smad3 from the laboratory of Peter ten Dijke (1 in 10 000), 3F10 rat monoclonal anti-HA from Roche (1 in 500) and SC-9141 rabbit polyclonal anti-SnoN from Santa Cruz Biotechnology (1 in 2000). Following overnight incubation, primary antibodies were removed and membranes were subjected to four 15-minute washes in PBST with gentle rocking. Secondary antibodies used as appropriate were polyclonal goat anti-mouse IgG conjugated to horseradish peroxidase (HRP), polyclonal goat anti-rabbit IgG conjugated to HRP and polyclonal rabbit anti-rat IgG conjugated to HRP (all from Dako). These were diluted 1 in 2000 in blocking solution and were incubated with the washed membranes for 1-3 hours at room temperature with gentle rocking. Membranes were again subjected to four 15-minute washes in PBST with gentle rocking, followed by several washes in PBSA to remove the Tween. Membranes were then briefly incubated with ECL immunoblot detection reagent (Amersham Pharmacia Biotech), dried and visualised by autoradiography with blue sensitive autoradiography film (GRI-Kodak) and a $Compact_2$ autoradiogram developer (IGP).

2.16 Silver staining of SDS-PAGE gels

Silver staining was carried out by the standard EMBL protocol (Dorey, K; personal communication). Following electrophoresis, gels were fixed for 20 minutes in 5 % acetic acid (v/v) and 50 % methanol (v/v) with gentle rocking. This was followed by washing for 10 minutes in 50 % methanol (v/v) with gentle rocking and further washing in water for 2 hours to overnight. Gels were then sensitised for 1 minute in 0.02 % sodium thiosulphate (w/v) with gentle rocking and subjected to two 1-minute rinses in water. This was followed by incubation with 0.1 % silver nitrate (w/v) for at least 20 minutes at 4 °C and in the dark. Gels were then rinsed for 1 minute with water and transferred to a fresh gel tray. Finally, development was carried out with 4 % sodium carbonate (w/v) and 0.04% Formalin (v/v) (added directly before use) until protein bands were of an appropriate intensity. The development reaction was stopped with 5 % acetic acid (v/v). Cellulose acetate sheets (Promega) were then immersed in 1 % glycerol (v/v) and the gel was clamped between these sheets in a plastic frame (Promega) until it had dried.

2.17 Synthesis of probes for electrophoretic mobility shift assays (EMSA)

EMSA probes were synthesised by the method of Germain *et. al.* (Germain et al., 2000). Pairs of partially overlapping antiparalell oligonucleotides corresponding to the DNA binding sites of interest were synthesised. The single-stranded overhangs produced upon hybridisation of these oligonucleotide pairs were then filled-in with *Taq* polymerase in the presence of radiolabelled dNTPs. Oligonucleotide sequences used are as shown in table 2.1. 10 µl probe synthesis reactions contained 1 µl of each oligonucleotide (50 ng μl^{-1}), 1 µl PCR buffer, 1 µl α^{32}P-dATP, 1 µl α^{32}P-dCTP, 1 µl dGTP/dTTP mix (0.5 mM each), 1 µl dATP/dCTP mix (50 µM each), 3 µl water and 1 µl *Taq* polymerase (Developmental Signalling Laboratory, Cancer Research UK). PCR buffer consisted of 500 mM KCl, 100 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$ and 0.01 % (w/v) gelatin. After synthesis probes were resolved on acrylamide gels having the same composition as in 2.17 below but poured with 0.4 mm spacers and

Table 2.1 Overlapping oligonucleotides used in the synthesis of EMSA probes

Probe	Sense	Antisense	Fig.
c-Jun SBR "full-length" (55bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.2 3.3
c-Jun SBR original truncation (35 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	GCTGTGTCTGTCTGTCTGC	3.3
c-Jun SBR upstream truncation 1 (50 bp)	TGCGCGGAGTCAGGCAGAC AGACAGACACAGC	TGCCGACCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.7
c-Jun SBR upstream truncation 2 (45 bp)	GGAGTCAGGCAGACAGACA GACACAGC	TGCCGACCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.7
c-Jun SBR upstream truncation 3 (40 bp)	CAGGCAGACAGACAGACAC AGC	TGCCGACCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.7
c-Jun SBR upstream truncation 4 (37 bp)	GCAGACAGACAGACACAGC	TGCCGACCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.7
c-Jun SBR downstream truncation 1 (53 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	CCGACCTGGCTGGCTGGCT GTGTCTGTCTGTCTG	3.8
c-Jun SBR downstream truncation 2 (51 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	GACCTGGCTGGCTGGCTGT GTCTGTCTGTCTG	3.8
c-Jun SBR downstream truncation 3 (49 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	CCTGGCTGGCTGGCTGTGT CTGTCTGTCTG	3.8
c-Jun SBR downstream truncation 4 (47 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGGCTGGCTGGCTGTGTCT GTCTGTCTG	3.8
c-Jun SBR "minimal" (32 bp)	CAGGCAGACAGACAGACAC AGCCAGCCAGCCA	TGGCTGGCTGGCTGTGT	3.9
c-Jun SBR mutant 1 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGGCTGGCATG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 2 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGGCATGCTGG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 3 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCATGCTGGCTGG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 4 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGGCTGATTGG CTGTGTCTGTCTGTCT	3.10 3.11
c-Jun SBR mutant 5 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGATTGGCTGG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 6 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGATTTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 7 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCTCCCTGGCTGGCTGG CTGTGTCTGTCTGTCTG	3.10 3.11
c-Jun SBR mutant 1+2 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCTGGCATGCATG CTGTGTCTGTCTGTCTG	3.12
c-Jun SBR mutant 1+3 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCATGCTGGCATG CTGTGTCTGTCTGTCTG	3.12
c-Jun SBR mutant 2+3 (55 bp)	GGAGGTGCGCGGAGTCAGG CAGACAGACAGACACAGC	TGCCGACCATGCATGCTGG CTGTGTCTGTCTGTCTG	3.12

comb. Gels were run for 1 hour at 150 V and 150 mA in 0.5X TBE buffer (see 2.19 below). Gels were then visualised 1 minute exposure of blue sensitive autoradiography film (GRI-Kodak) and development on a $Compact_2$ autoradiogram developed (IGP). The film was used to identify the position of probe bands on the gel. These were then excised and immersed overnight at 37 °C in 350 µl of 0.3 M sodium acetate, 0.1 % (w/v) SDS and 1 mM EDTA. The supernatant was then taken and mixed with 350 µl phenol before centrifugation for 2 minutes at 12 500 rpm on a Biofuge Pico centrifuge (Sorvall). After this, the upper aqueous phase was taken and probe was precipitated for 30 minutes with 1 µg tRNA and 700 µl 96 % ethanol on dry ice. Samples were then centrifuged for 20 minutes at 12 500 rpm and 4 °C on a Biofuge Fresco Heraeus centrifuge (Sorvall) and the supernatant was removed. The pelleted probe was washed with 70 % ethanol and centrifuged again for 5 minutes at 12 500 rpm and 4 °C on a Biofuge Fresco Heraeus centrifuge (Sorvall). The supernatant was removed and the pellet was dried for 3-5 minutes on a DNA 110 speed vac dessication system (Savant). Finally, the dry pellet was resuspended in 20 µl TE buffer (10 mM Tris-HCl [pH 7.5] and 1 mM EDTA [pH 8.0]).

2.18 Quantification of probes for EMSA

The following general procedure was applied in calculating the concentration of radiolabelled probes. It is based on the method of Pollock and Treisman (Pollock and Treisman, 1990). 1 µl samples of each probe were taken and the number of radioactive decays occurring per minute was determined using an LS 65 000 scintillation counter (Beckman). Samples were then returned to the appropriate stock of probe. Division of the number of counts per minute (CPM) by 10^6 and multiplication by 20 (the ratio of the volume of probe stock to the sample quantified) yields the activity of the probe in Ci. Division of this value by the mass specific activity of the probe (in Ci ng^{-1}) and multiplication by the radioactive decay factor for ^{32}P ($e^{-0.0495t}$, where t is the number of days since the reference date on the stock of radiolabelled dATP and dCTP) yields the quantity of probe in ng. The mass specific activity of each probe was determined in the following way. First the molar ratio of labelled to unbalelled dATP and dCTP was calculated. Then the number of positions

where an A or C base is incorporated during probe synthesis was counted. Multiplication of these two quantities yields the number of radiolabelled nucleotides incorporated into each probe. Multiplication of this by the molar activity of the labelled dNTPs at their reference date (3×10^6 Ci mol^{-1}) yields the molar activity of the probe in Ci mol^{-1}. Finally, division of this value by the molecular mass of the double-stranded probe (in ng mol^{-1}) yields the mass specific activity of the probe. Once the quantity of probe present in a 20 µl volume of TE buffer is known, the volume is adjusted with additional TE to bring the concentration down to 0.2 ng μl^{-1}.

2.19 Electrophoretic mobility shift assays

To analyse protein-DNA interactions, EMSA was carried out by the method of Germain *et. al.* (Germain et al., 2000). Nuclear extracts were adjusted to a protein concentration of 1 µg μl^{-1} by dilution in hypertonic buffer and 10 µl samples were taken and incubated for 5 minutes with 1 µl volumes of appropriate antibodies and peptides. Concentrations of wild-type and mutant SIM peptide used in Fig. 3.3A were 25, 50, 100, 150, 200, 400 and 500 pmol μl^{-1}. Concentrations used in Fig. 3.3B were 25, 50, 150, 250 and 500 pmol μl^{-1}. In all other figures, 500 pmol μl^{-1} were used. The sequence of the wild-type SIM peptide was biotin-aminohexanoic acid-RQIKIWFQNRRMKWKKLLMDFNNFPPNKTITPDMNVRIPPI. The sequence of the mutant SIM peptide was biotin-aminohexanoic acid-RQIKIWFQNRRMKWKKL LMDFNNFAAAKTITPDMNVRIPPI. In samples where fewer antibodies/peptides were added, the additional volume was made up with water. 10 µl of probe mix were then added. This consisted of 1 µl 200 mM KCl, 1 µl 110 mM $MgCl_2$, µl $ZnSO_4$, 1 µl probe (at 0.2 ng μl^{-1}), 1 µl poly-deoxyinosinate-deoxycytidine (at 1 µg μl^{-1}), 4 µl 50 % (v/v) glycerol and 1 µl water. Samples were then briefly centrifuged in a Biofuge Pico centrifuge (Sorvall) and incubated at room temperature for 10 minutes before loading on a bandshift gel and electrophoresis for 3-4 hours at 150 V and 150 mA. Gels were prepared by placing 1.5 mm plastic spacers between two glass plates (at opposite edges of the plates) and sealing three edges of the plates with electrical tape, leaving one edge open for gel mix. 50 ml of gel mix consisting of 5 % acrylamide, 0.125 % bisacrylamide, 0.5X TBE buffer and 2.5 % (v/v) glycerol were prepared.

125 µl of each of TEMED and ammonium persulphate were added, the mix was poured into the space between the glass plates and a comb bearing an appropriate number of teeth was inserted. Gels were then allowed to polymerise for 15 minutes before removal of the tape and comb and clamping into an electrophoresis tank (Cambridge Electrophoresis). 0.5X TBE buffer (0.045 M Tris borate and 0.5 mM EDTA at pH 8.0) was then added before loading of samples. Following electrophoresis gels were transferred to 3MM Whatman paper and dried for 45 minutes at 80 °C on a Hoefer Drygel Sr. vacuum-driven gel dryer (Pharmacia Biotech). Dried gels were then placed in an autoradiography case with Hyperfilm autoradiography film (Amersham Biosciences) and placed at -80 °C or room temperature overnight or for several days. Autoradiograms were developed using a Compact$_2$ autoradiogram developer (IGP).

2.20 Preparation of *in vitro* translated proteins

^{35}S-labelled protein was prepared using the TNT Coupled Reticulocyte Lysate System (Promega) as per the manufacturer's instructions. Samples were taken, heated with SDS sample buffer at 95 °C and run on an SDS-PAGE gel. Gels were then dried and visualised by autoradiography at room temperature as described in 2.18 above to check the integrity and yield of synthesised proteins.

2.21 McKay Assays

Radiolabelled probes were synthesised and quantified as in 2.17 and 2.18 above. Probe sequences used were CCGACTAGTATCTGCTGCCCTAAAATGTGTATTC CATGGAAATGTCAGCCCTTCTCTCCCTAGCTAGCCGG (ARE), GTACGCGA GTGTGTATTGGAACGTAACTCATT (FoxH1a selected probe) and TGGGTGTG TATTGGGATGTGTATTGAGCGCAGG (FoxH1b selected probe). McKay assays were carried out essentially as described by Pollock and Treisman (Pollock and Treisman, 1990). 2 ml of *in vitro* translated protein obtained as in 2.20 above were combined with 2 ml of water and 18 ml of buffer containing 9 mM HEPES pH 7.5, 11

mM HEPES pH 7.9, 10 % (v/v) glycerol, 0.18 M KCl, 0.18 % (v/v) triton X-100, 5.6 mM EGTA and 2.8 mM EDTA, 1.4 mM $MgCl_2$, 0.06 % (v/v) NP-40, 0.9 mM dithiothreitol and 28 $\mu g\ ml^{-1}$ bovine serum albumin. After this, 1 μl 200 ng μl^{-1} poly-deoxyinosinate-deoxycytidine and 1 μl radiolabelled probe (at 0.2 ng μl^{-1}) were added and reactions were incubated at room temperature for 30 minutes. M2 anti-Flag–ProteinA–Agarose beads (Sigma-Aldrich) were washed three times in the above buffer diluted by a factor of 1.22 in water. All free liquid was then removed from the beads and the above reactions were added to the beads. It is important that this was carried out in a 0.5 ml tube. These reactions were then incubated for 2 hours at 4 °C with gentle agitation before scintillation counting, removal of the supernatant, washing and elution. Three sequential washes were carried out with a buffer consisting of 20 mM HEPES pH 7.9, 200 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 2.5 mM $MgCl_2$, 0.1 % (v/v) NP-40 and 2 mM dithiothreitol. Elution was carried out by heating for 5 minutes at 95 °C with SDS sample buffer. Whole eluates were subjected to scintillation counting and the proportion of radiolabelled probe pulled-down was calculated. Eluates were also analysed by immumoblotting as described above to determine the quantity of *in vitro* translated protein pulled down.

2.22 Initial synthesis of probes for *in vitro* selection of transcription factor binding sites

Probes were synthesised by the method of Pollock and Treisman (Pollock and Treisman, 1990). Single-stranded full-length template for the synthesis of site selection probe was GCTGCAGTTGCACTGAATTCGCCTCGGAGTCAGGCAGA CAGACAGACNNNNNNNNNNNNNNNNNNNGACAGGATCCGCTGAACTG ACCTG and primers used for amplification were CAGGTCAGTTCAGCGG ATCCTGTCG (reverse) and GCTGCAGTTGCACTGAATTCGCCTC (forward). Initial probe synthesis reactions consisted of 2 μl full-length template (50 ng μl^{-1}), 2 μl of reverse primer (80 ng μl^{-1}), 2 μl PCR buffer (as described in 2.16 above), 2 μl dATP/dGTP/dTTP mix (each at 0.5 mM), 2 μl dCTP (40 μM), 2 μl α^{32}P-dCTP, 1 μl *Taq* polymerase (Developmental Signalling Laboratory, Cancer Research UK) and 4 μl water. Synthesis reactions were carried out in a GeneAmp PCR System 9700

(Applied Biosystems) with the following thermal cycle: 30 seconds at 94 °C, 3 minutes at 37 °C and 10 minutes at 72 °C. 2 µl of 0.5 mM unlabelled dCTP were then added to ensure complete filling-out and reactions were heated for a further 10 minutes at 72 °C in the thermal cycler. Probes were purified as in 2.17 above (except with 10 µg glycogen in place of 1 µg tRNA in the precipitation step) and quantified as in 2.18 above.

2.23 *In vitro* selection of transcription factor binding sites

Site selection was carried out by the method described by Pollock and Treisman (Pollock and Treisman, 1990). Antibodies specific to Smad2/Smad3 (BD Bioscience) and Smad3 (Zymed) were bound to protein A/G beads as described in 2.11 above. Binding reactions were then set up as follows in 0.5 ml tubes: 20 µl nuclear extract at a concentration of 1 µg μl^{-1}, 51 µl hypotonic buffer as described in 2.8 above, 3 µl radiolabelled probe (at 0.2 ng μl^{-1}) and 1.5 µl poly-deoxyinosinate-deoxycytidine (200 ng μl^{-1}). These were incubated for 30 minutes at room temperature before addition of 20 µl bead slurry. Reactions were then incubated at 4 °C for 2 hours with gentle agitation. The supernatant was removed and beads were washed three times with hypotonic buffer containing 140 mM NaCl. Probe was then eluted by incubation for 1 hour at 45 °C in 50 mM Tris-HCl (pH 8.0), 100 mM sodium acetate, 5 mM EDTA and 0.5 % SDS. Eluted probe was precipitated and quantified as in 2.21 and amplified by PCR. The following PCR reactions mixtures were used: 2 µl selected probe (50 ng μl^{-1}), 2 µl of each primer (80 ng μl^{-1}), 2 µl PCR buffer (as described in 2.16 above), 3.2 µl dATP/dGTP/dTTP mix (each at 0.5 mM), 2 µl dCTP (40 µM), 1 µl α^{32}P-dCTP, 1 µl *Taq* polymerase (Developmental Signalling Laboratory, Cancer Research UK) and 6 µl water. These reactions were then subjected to the following thermal cycle: 1 minute at 94 °C; 19 cycles of 30 seconds at 94 °C, 30 seconds at 37 °C and 30 seconds at 72 °C; 2 minutes at 72 °C and an indefinite time at 10 °C. PCR products were extracted and quantified as in 2.22 above. This cycle of immunoprecipitation followed by PCR was repeated four times before analysis of complex formation on the final probe by EMSA as in 2.19 above.

2.24 Protein-DNA crosslinking by ultraviolet irradiation

Crosslinking and analysis were carried out by a method previously used by Graham Esslemont (Esslemont, GM and Hill, CS; unpublished data). Probes were synthesised as in 2.17 above but with BrdUTP substituted in place of dTTP. Probes were then quantified as in 2.18. EMSA was carried out as in 2.19 above using combs with 1 cm teeth. One of the glass plates was then removed from the gel and the gel was covered with Saran polyvinyldiene chloride (PVDC) wrap (Dow chemical company). The gel was then placed face-down on an ultraviolet transilluminator (UVP) and a plastic box with a greater surface area than the gel and containing ice was placed over it. The transilluminator was then set to 245 nm and switched on. The gel was then immersed for 15 minutes in a buffer containing 44.5 mM Tris.HCl pH8 and 1.25 mM EDTA. Cellulose acetate sheets (Promega) were then immersed in 1 % glycerol (v/v) and the gel was clamped between these sheets in a plastic frame (Promega) until it had dried. Gels were then visualised by autoradiography as described in 2.18 above and gel slices corresponding to protein-DNA complexes of interest were excised. Gel slices were rehydrated by immersion for 15 minutes in a solution containing 50 mM Tris-HCl pH6.8 and 1% ß-mercaptoethanol. Cellulose acetate sheets were then removed from either side of the gel slices and the slices were immersed for 20 minutes in a solution containing 50 mM Tris.HCl pH6.8, 1% ß-mercaptoethanol and 0.1% SDS. Gel slices were then loaded into the wells of an SDS-PAGE gel (see 2.14 above) poured with 1.5 mm spacers and 1.5 cm wells and the following solution was injected in molten form to create a seal: 120 mM Tris-HCl pH6.8, 1% agarose, 0.1% SDS and a small quantity of bromophenol blue. The agarose was allowed to set and gels were run as described in 2.14 above. Gels were then visualized by autoradiography as above.

TGF-β-DEPENDENT COMPLEX FORMATION ON THE *c-Jun* PROMOTER

3.1 Introduction

In earlier studies, Inman and Hill investigated a TGF-β-dependent complex that is assembled on the *c-Jun* Smad-binding region (SBR) *in vitro*. This complex was found to consist of one molecule of each of Smad3 and Smad4 plus two unidentified factors. One of these factors – factor Y – is thought to make direct contact with DNA while the other –factor X – appears to bind Smad3 through a sequence functionally equivalent to the Mixer SIM and probably does not contact DNA directly (Fig. 1.9B) (Inman and Hill, 2002).

These findings provided an exciting opportunity to study Smad-transcription factor interactions in fully differentiated cells of the mammalian adult. While several transcription factors have been reliably shown to cooperate with Smads in this context (see Chapter 1 above), none have been characterised in the same molecular detail as members of the FoxH1 and Mixer families in *Xenopus* embryos. The discovery that cells of the mammalian adult harbour factors containing sequences functionally equivalent to the SIM (Inman and Hill, 2002) provides an ideal starting point for biochemical investigation of Smad-transcription factor interactions in this setting.

The work of Inman and Hill (Inman and Hill, 2002) also provides an opportunity to characterise physiological cooperation of Smad3 with transcription factors at the molecular level. Smad3 is not expressed in early *Xenopus* embryos and, while it can interact with members of the FoxH1 and Mixer families, these interactions are not physiological (Randall et al., 2004). It is true that most of the Smad-transcription factor interactions that have been characterised in adult tissues are with Smad3 but these interactions have not been thoroughly investigated on the molecular level (see above). A chance thus exists to achieve a deeper understanding of the interaction of Smad3 with transcription factors.

The first objective of my work was to confirm the existence of additional factors in the Smad-containing complex that forms on the *c-Jun* SBR in response to TGF-β and this is the central theme of the present chapter. Confirming the presence of these additional components is a necessary prerequisite for their subsequent identification and molecular characterisation. A combination of electrophoretic mobility shift assay (EMSA) and various pull-down experiments were used to investigate the protein composition of the complex under various conditions. Following this, probe truncation and mutagenesis were used to map the minimal *c-Jun* SBR and to determine the precise binding site for factor Y. *In vitro* binding site selection was also attempted to determine the physically optimal binding site for factor Y. Taken together, these results show conclusively that factors X and Y are indeed present in the complex.

3.2 The Smad3-Smad4 interaction is stabilised on the *c-Jun* SBR

My initial studies centred on investigation of the protein composition of the complex under different conditions and uncovered evidence that the complex is assembled on DNA in a TGF-β-dependent fashion – as opposed to being formed in the nucleoplasm of TGF-β-induced cells and then binding DNA as a pre-assembled entity. The work of Inman and Hill had already shown that the interaction of Smad3 and Smad4 with the *c-Jun* Smad-binding region (SBR) is dependent on factors X and Y and that factor Y cannot bind DNA in the absence of Smads binding to an adjacent site (Inman and Hill, 2002). These findings support a model in which protein complex formation occurs largely on the SBR, with new interactions being established or weak affinities being potentiated by the DNA scaffold.

Since it was thought that complex formation occurs largely on DNA, I thought it likely that the interaction between Smad3 and Smad4 is also potentiated on the *c-Jun* SBR. This hypothesis was tested by immunoprecipitation carried out in the presence of double-stranded wild-type oligonucleotide corresponding to the *c-Jun* SBR or mutant oligonucleotide in which the SBE and factor Y binding site are disrupted (Fig. 3.1A). This was carried out using nuclear extract from TGF-β-induced or uninduced

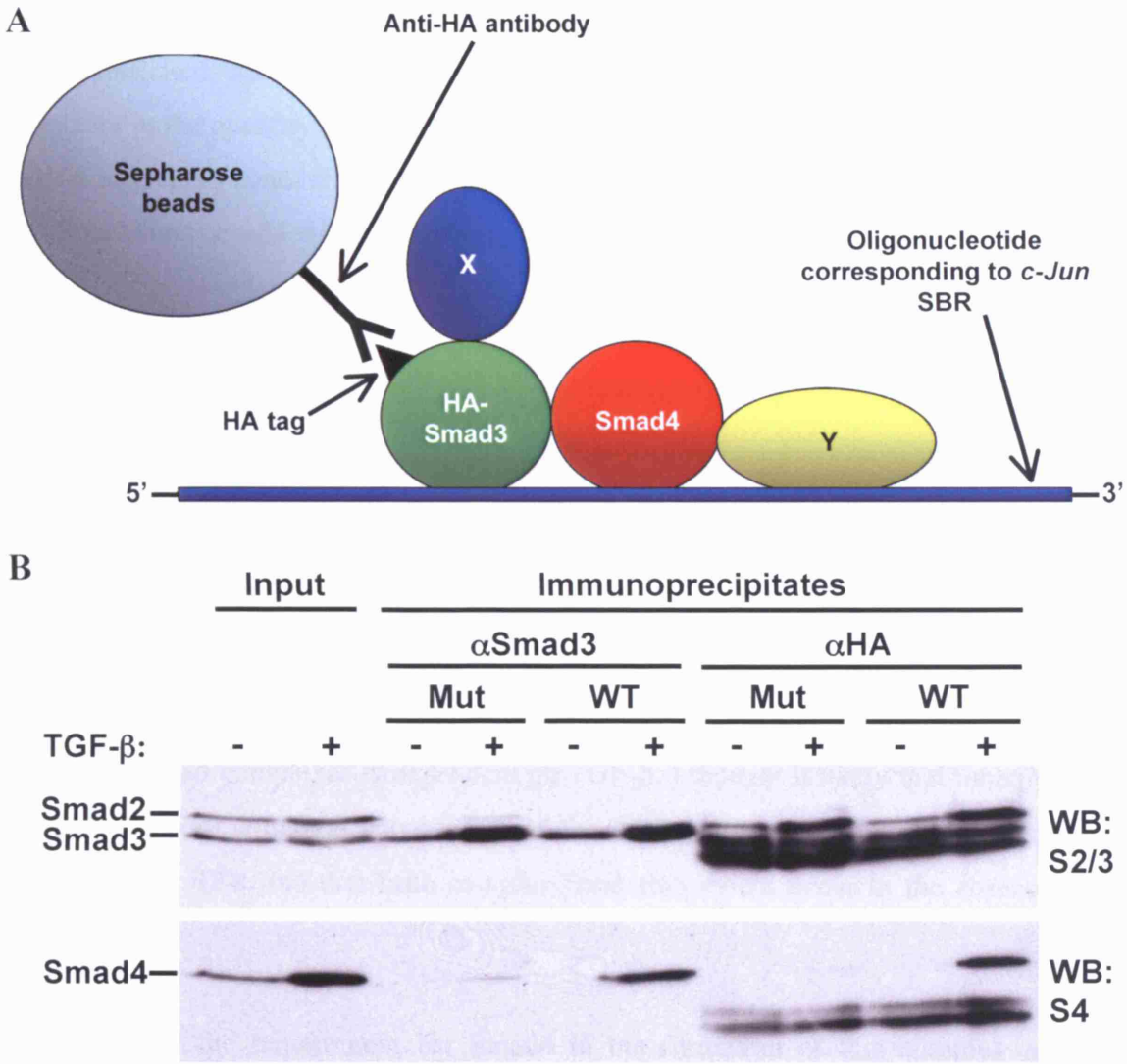


Fig. 3.1 The interaction between Smad3 and Smad4 is stabilised on the *c-Jun* Smad-binding region (SBR). *A) Schematic showing immunoprecipitation (IP) of haemaglutinin (HA)-tagged Smad3 in the presence of an oligonucleotide corresponding to the* c-Jun *SBR.* Incubation is carried out as above with the protein components of the complex supplied by HaCaT nuclear extract. Beads were then washed and boiled in SDS sample buffer. The same procedure was also carried out with an antibody specific to endogenous Smad3. *B) Immunoprecipitation as described in A.* Nuclear extract from HaCaT cells stably expressing HA-Smad3 was immunoprecipitated with anti-HA or anti-Smad3 antibodies, as indicated. Immunoprecipitates and inputs were analysed by western blotting (WB) with antibodies specific to Smad2/Smad3 (S2/3) and Smad4 (S4). Inputs run on gel were 10% total protein compared to immunoprecipitation inputs.

HaCaT cells stably expressing HA-Smad3. Antibodies specific to Smad3 or the HA tag were used and wild-type or mutant oligonucleotides were included. The results showed that, under conditions where DNA-independent co-precipitation of Smad4 was inefficient, addition of wild-type but not mutant oligonucleotide yielded a large increase in the quantity of Smad4 that was co-precipitated (Fig. 3.1B). This indicates DNA sequences containing SBE sites can act as a scaffold to stabilise the interaction of Smad3 and Smad4 with each other,

3.3 Smad4 is required for TGF-β-dependent complex formation on the *c-Jun* SBR

As a general principle, the formation of transcription factor complexes on gene promoters and enhancers tends to be dependent on multiple distinct components of the complex – if any of these components are not available, complex formation does not occur and the gene is not induced (Fig 1.1) (Courey, 2001). Since the formation of active Smad complexes is dependent on TGF-β, I thought it likely that Smad3 and Smad4 are core structural components of the TGF-β-dependent complex that forms on the *c-Jun* SBR and that little complex formation would occur in the absence of either factor.

I thus tested the requirement for Smad4 in the formation of this complex using HaCaT-TRS4 cells, a HaCaT cell line expressing an shRNA against Smad4 in a tetracycline-inducible fashion (Levy and Hill, 2005). Immunoblot analysis shows that, in the absence of tetracycline (Tet), these cells express a lower level of Smad4 than wild-type HaCaTs and that, when Tet is added, expression becomes lower still (Levy and Hill, 2005). EMSA analysis of wild-type HaCaT cells and HaCaT-TRS4 cells treated with tetracycline (Tet) or left untreated showed that TGF-β-induced complex formation on the *c-Jun* SBR diminishes in proportion to Smad4 expression levels. Levels of TGF-β-dependent complex formation are high in wild-type HaCaTs, lower in HaCaT-TRS4 cells not treated with Tet and lower still in Tet-treated HaCaT-TRS4 cells (Fig. 3.2). These data indicate that Smad4 is required for complex formation.

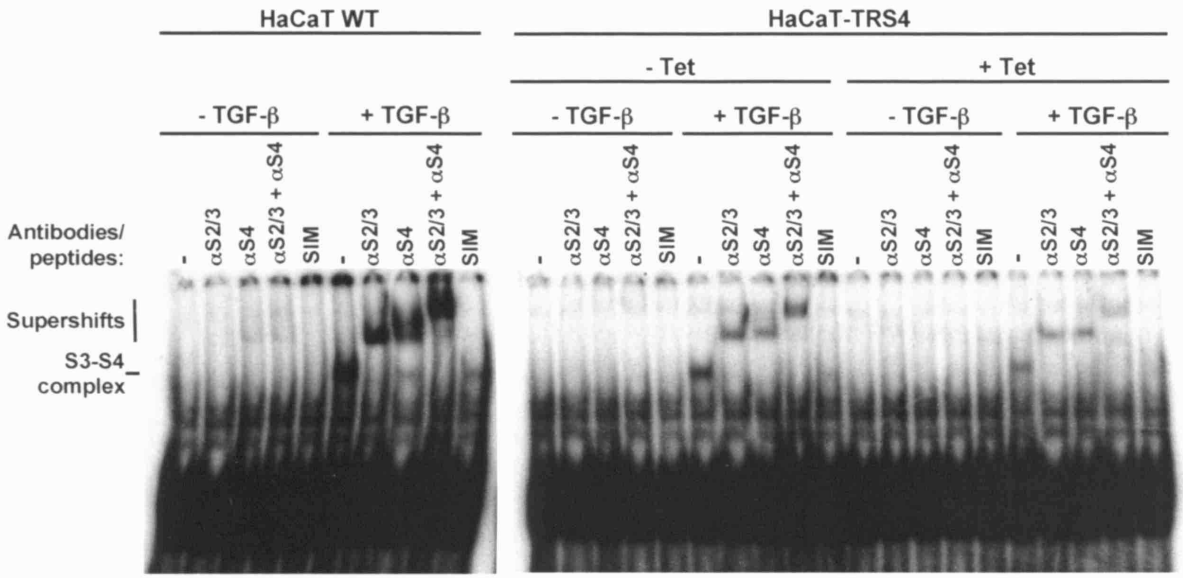


Fig. 3.2 Smad4 is limiting for TGF-β-inducible complex formation on the *c-Jun* SBR. An electrophoretic mobility shift assay (EMSA) was carried out using a radiolabelled oligonucleotide corresponding to the *c-Jun* SBR and nuclear extract from wild-type (WT) HaCaT cells and HaCaT-TRS4 cells, which express an shRNA against Smad4 in a tetracycline (Tet)-inducible manner (Levy and Hill, 2005). HaCaT-TRS4 cells were treated with tetracycline (Tet) or left untreated. Both cell lines were treated with TGF-β for 1h or left untreated. Supershifts were carried out as indicated using antibodies that recognise Smad2/Smad3 (αS2/3) and Smad4 (αS4) . Complexes were disrupted as indicated with a peptide corresponding to the Mixer SIM (lanes marked "SIM"). The positions of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated.

3.4 Confirming the existence of factors X and Y

Having investigated the interaction between Smad3 and Smad4 on the *c-Jun* SBR and having shown that Smad4 is limiting for complex formation, it was necessary to confirm the presence of factors X and Y in the complex. I achieved this mainly by EMSA using a radiolabelled probe corresponding to the *c-Jun* SBR. In these analyses, a specific complex bound to the probe in a TGF-β-inducible fashion and was supershifted by antibodies specific to Smad2/Smad3 and Smad4 – consistent with the presence of Smad3 and Smad4 in the complex (Fig. 3.3A). Increasing concentrations of SIM peptide were also added to competitively disrupt the complex. It was found that this caused both a dose-dependent loss of complex stability and an increase in electrophoretic mobility (Fig. 3.3A). Both these effects continued to a certain plateau level and, beyond this point, were not enhanced by further increases in SIM concentration. This probably reflects the SIM competition achieving complete disruption of the Smad3–factor X interaction, making it impossible for further excess of SIM to have an enhanced competitive effect. These results reproduce the findings of Inman and Hill (Inman and Hill, 2002) and support the existence of factor X.

In these same studies, EMSA was also carried out with a radiolabelled probe in which the *c-Jun* SBR and downstream residues were truncated. This resulted in a dramatic loss of complex stability (Fig. 3.3B), which is consistent with the findings of Inman and Hill (Inman and Hill, 2002) and with the existence of factor Y. Addition of SIM peptide resulted in a further loss of complex stability and higher concentrations of SIM eliminated all detectable complex (Fig3.3B).

There are two possible interpretations of these findings: that factors X and Y are in fact distinct proteins or that they represent separate binding activities residing in different domains of the same polypeptide. The latter would in some ways be more consistent with what is known about the FoxH1 and Mixer families. These are DNA-binding factors that also bind Smad2 via a SIM motif and which are known to be able to associate with Smad2-Smad4 complexes in the absence of their cognate DNA binding sites – which would be similar to what happens on the truncated *c-Jun* SBR probe if factors X and Y are in fact the same molecule. However, Mixer and FoxH1

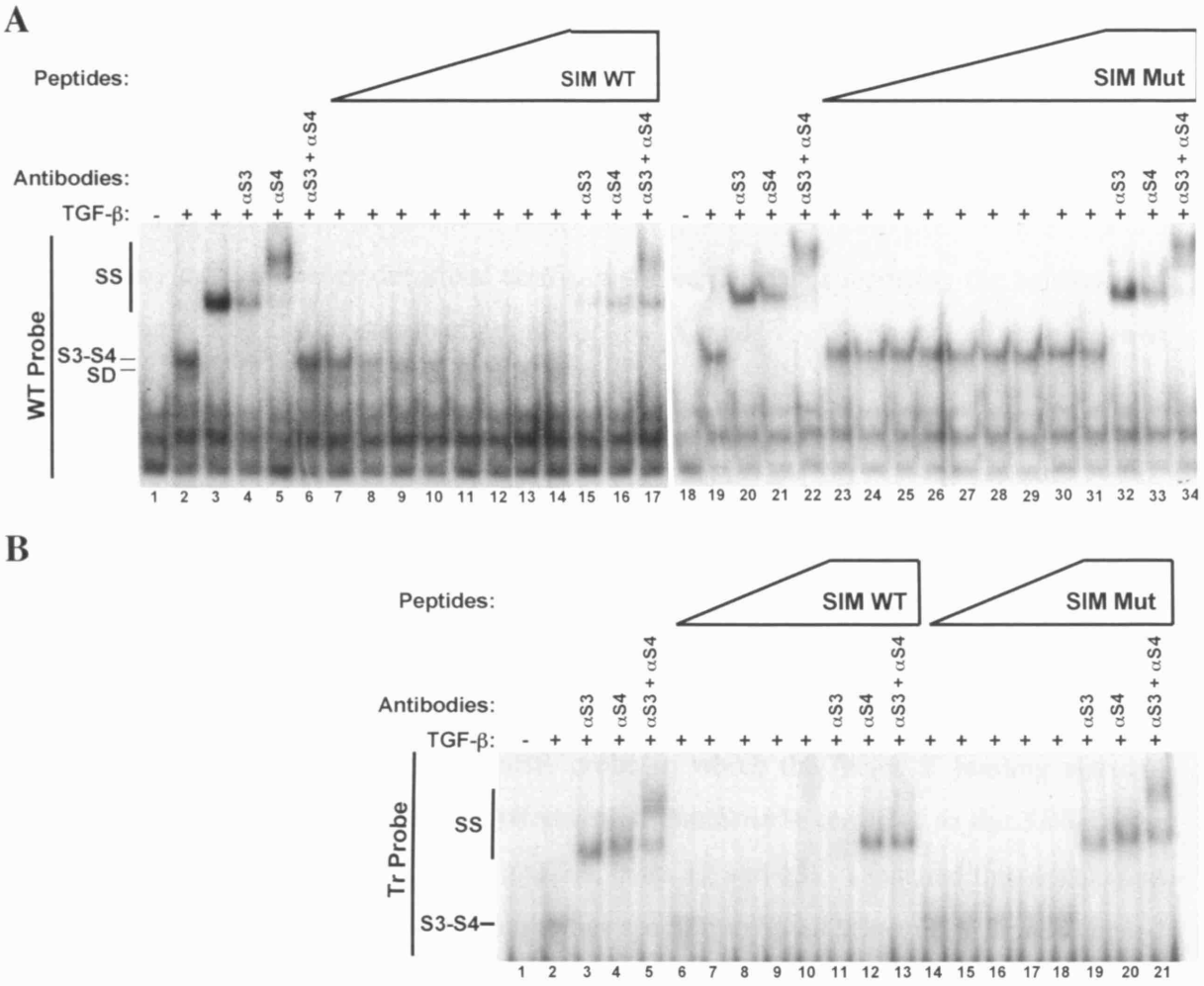


Fig. 3.3 Smad4 can bind the *c-Jun* SBR without associated Smad3. *A) SIM peptide disrupts TGF-β-dependent complex formation on the* c-Jun *SBR.* EMSA using nuclear extract from TGF-β -induced or uninduced HaCaT cells. Radiolabelled oligonucleotide corresponding to the *c-Jun* SBR was used as wild-type (WT) probe. Antibodies specific to Smad2/Smad3 (αS2/3) and to Smad4 (αS4) were added to indicated lanes, as was wild-type (WT) or mutant (Mut) peptide corresponding to the Mixer SIM. The positions of the TGF-β-induced Smad3-Smad4 complex (S3-S4), of antibody-induced supershifts of this complex (SS) and of the SIM-disrupted complex (SD) are indicated. Note that, as well as affecting complex stability, SIM competition causes a slight increase in electrophoretic mobility. This is best seen by using the middle band of the triplet of background bands present on the gel as a reference point. The distance between this band and the TGF-β-dependent complex is greater in lane 6 than in lane 14. *B) SIM peptide competition and deletion of the factor Y binding site have an additive effect on complex stability but still allow formation of a complex containing Smad4 but not Smad3.* EMSA carried out exactly as above but using probe in which the factor Y binding site has been truncated (Tr) (Inman and Hill, 2002). Refer to Fig. 1.9B for the position of the factor Y binding site in the WT probe.

make the major contribution to DNA binding in the Smad-containing complexes in which they are involved. This is unlike factor Y, which appears to bind DNA only after Smads have bound to an adjacent site (Inman and Hill, 2002). In addition, transcriptional regulatory complexes generally consist of a large number of distinct subunits and it thus appears more likely that factors X and Y are in fact distinct proteins.

In any case, the work described confirms the earlier data supporting the existence of binding activities corresponding to factors X and Y. Further confirmation was subsequently provided by mutational analysis of the factor Y binding site and pull-down assays that indicate the presence of at least one non-Smad factor in the complex (see below).

3.5 Smad4 can bind the *c-Jun* SBR in the absence of Smad3

EMSA carried out with *c-Jun* SBR probe in which the factor Y binding site and downstream residues are truncated revealed that Smad4 can bind to the SBR in the absence of associated Smad3 (Fig. 3.3B, lanes 12 and 13). I reached this conclusion on the basis of supershifts with antibodies specific to Smad2/Smad3 or Smad4 carried out in the presence of high concentrations of SIM peptide. Under these conditions anti-Smad4 can stabilise and supershift an otherwise undetectable complex but anti-Smad2/3 cannot (Fig 3.3B). To confirm that this observation is not the result of anti-Smad4 binding directly to DNA, I carried out an EMSA with this antibody, truncated *c-Jun* SBR probe and extract from MDA-MB231 cells, which do not express Smad2, Smad3 or Smad4. This revealed no direct interaction between the anti-Smad4 antibody and DNA (data not shown). It can thus be concluded that, when high concentrations of SIM are added, a complex that contains Smad4 but not Smad3 forms on the truncated probe. The fact that the antibody specific to Smad2/Smad3 has also been shown to stabilise TGF-β-dependent complexes on the *c-Jun* SBR (Fig 3.3A) but does not stabilise any complex on truncated probe in the presence of SIM (Fig 3.3B) makes it very likely that Smad3 is genuinely absent from these complexes.

Figure 3.4 summarises the composition and stability of the complex under these *in vitro* manipulations. On "full-length" probe in the absence of SIM peptide, the complex consists of Smad3, Smad4, factor X and factor Y and is at maximal stability. When SIM peptide is added, factor X cannot bind to Smad3 and the complex is destabilised. Similarly, probe truncated downstream on the Smad binding site is used, the factor Y binding site is absent and so is factor Y. This also destabilises the complex. Finally, when truncated probe is used and SIM peptide is included, neither factor X, factor Y nor Smad3 are present in the complex. Under these conditions the complex is destabilised to the point where it can only be detected on EMSA when stabilised by anti-Smad4 antibody. The absence of Smad3 from these complexes is interesting and hints that factor Y may interact directly with Smad3, rather than Smad4.

3.6 In the absence of TGF-β, Smad4 binds the *c-Jun* SBR without Smad3

SIM competition studies reveal that Smad4 can associate with the *c-Jun* without associated Smad3, but this provides only the most tenuous indication that such complexes are relevant *in vivo*. However, it does provide a hint that they might have a biological function. Since Smad4 is present in the nucleus in the absence of TGF-β but active Smad3 is not, the most likely context in which a complex might be found *in vivo* is the TGF-β uninduced state. Such a complex could function to repress transcription, to nucleate assembly of the active complex or to keep the promoter region in a state where it is competent for activation – possibly by recruiting specialised histone modifying enzymes or chromatin remodelling complexes.

In order to test the hypothesis that Smad4 associates with the *c-Jun* promoter in the absence of TGF-β, I carried out DNA pull-down (DNAP) with wild-type oligonucleotide corresponding to the *c-Jun* SBR or mutant oligonucleotide with the SBE and factor Y binding sites scrambled. These oliognucleotides were 5'-biotinylated on one strand and immobilised on neutravidin beads. Nuclear extract was incubated with the beads, allowing the formation of TGF-β-dependent complexes

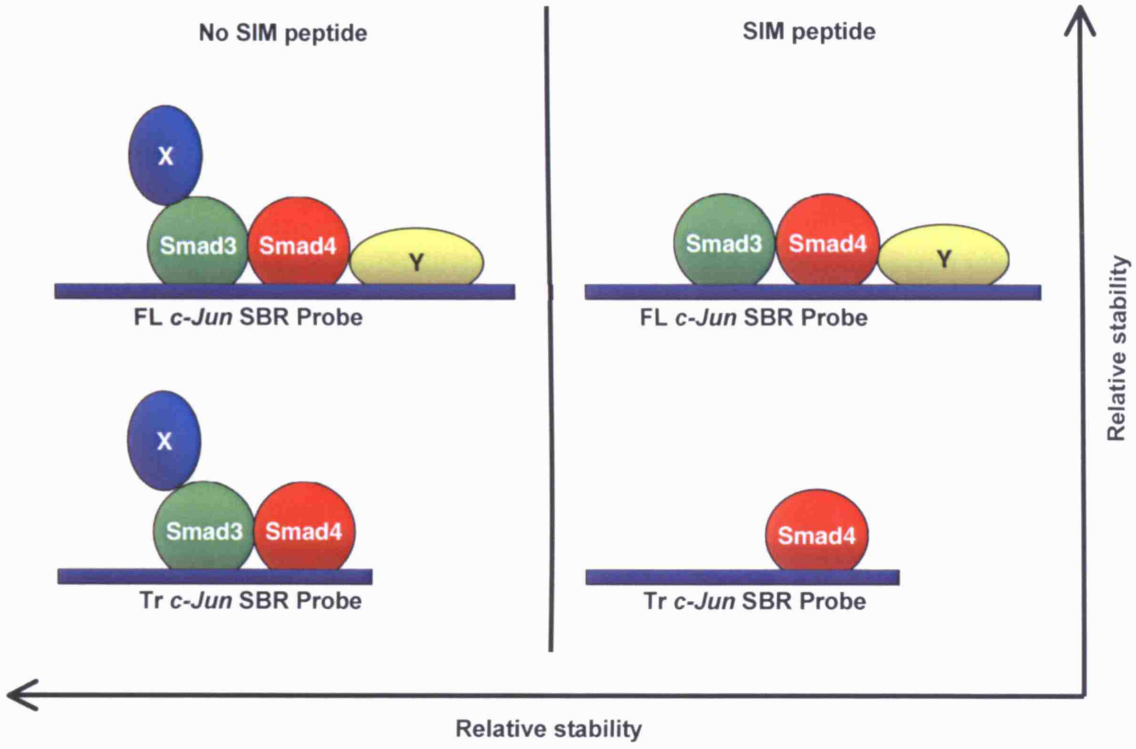


Fig. 3.4 States of the complex under *in vitro* manipulation. Compared to complexes formed on "full-length" (FL) *c-Jun* SBR probe, use of probe truncated in the factor Y binding site (Tr *c-Jun* SBR Probe) reduces complex stability by preventing the incorporation of factor Y into the complex. Addition of SIM peptide to the reaction also reduces complex stability - in this case by preventing the incorporation of factor X. Combination of both these manipulations reduces complex formation to very low levels and also prevents incorporation of Smad3.

on the oligonucleotide (Fig 3.5A). After this, the beads were removed, washed, boiled in SDS sample buffer and analysed by SDS-PAGE.

Subsequent immunoblot analysis revealed that, as predicted, Smad3 and Smad4 are pulled-down by wild-type but not mutant oligonucleotide in a TGF-β-dependent fashion (Fig. 3.5B). Small amounts of Smad2 were also pulled down in a specific fashion, suggesting that it can be incorporated into the complex at very sub-stoichiometric levels. In support of the model that Smad4 binds the *c-Jun* SBR in the absence of TGF-β, low levels of Smad4 were pulled down in this condition by wild-type but not mutant probe. Neither Smad2 nor Smad3 were pulled down under these conditions, consistent with the view that they do not associate with this element in the absence of TGF-β (Fig. 3.5B).

EMSA analysis also supported these conclusions. When carried out with full-length *c-Jun* SBR and nuclear extract from TGF-β-uninduced HaCaT cells I found that anti-Smad4 antibody could stabilise a low-abundance complex on this element but anti-Smad2/Smad3 could not. SIM peptide competition had no effect on this complex, reflecting the absence of Smad3 and factor X (Fig. 3.5C). It can thus be concluded that, in the absence of TGF-β, a complex forms on the *c-Jun* SBR that contains Smad4 but not Smad3. Possible functions of this complex are discussed in Chapter 5 below.

3.7 SnoN is also associated with the *c-Jun* SBR in the absence of TGF-β

If a complex containing Smad4 does exist on the *c-Jun* SBR *in vivo* in cells that have not been exposed to TGF-β, it is likely that other factors are also present. One possible function of such a complex would be to repress *c-Jun* expression in the absence of TGF-β. As a consequence, I thought it possible that the Smad corepressor SnoN might be a component of such a complex. This could be recruited by Smad4 or by other means and could antagonise transcription by interfering with the interaction between Smad3 and Smad4 or by recruiting corepressors such as N-CoR, with which SnoN is known to interact (Luo, 2004). Since SnoN is rapidly degraded in response

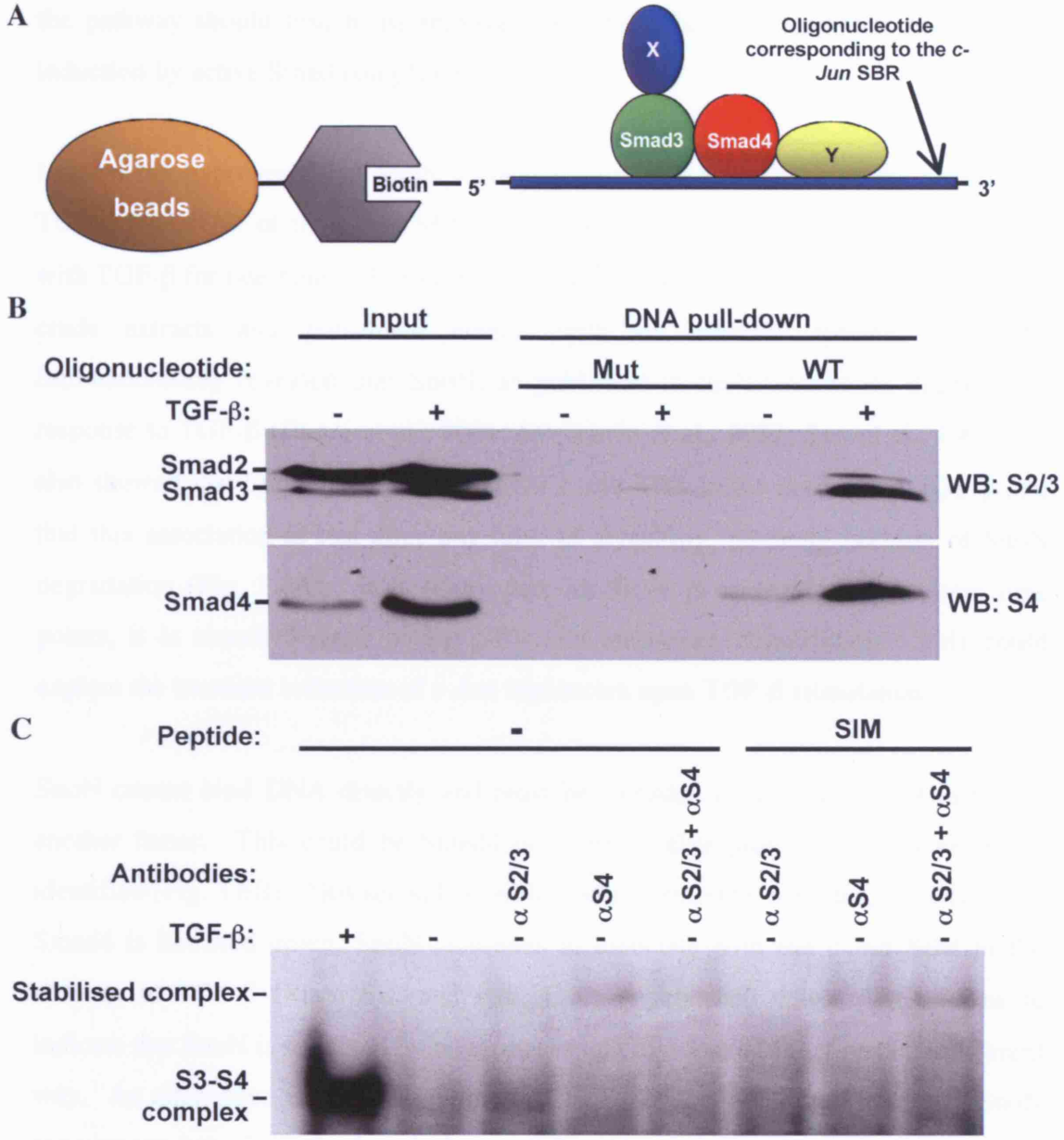


Fig. 3.5 Smad4 but not Smad3 is associated with the *c-Jun* SBR in the absence of TGF-β. *A) Schematic showing DNA pull-down (DNAP) of the* c-Jun *SBR and associated proteins. B) DNA pull-down of the* c-Jun *SBR.* Pull-downs were incubated as in A using oligonucleotide baits corresponding to wild-type and mutant *c-Jun* SBR, washed and boiled in SDS sample buffer. Precipitates were western blotted (WB) with antibodies specific to Smad2/Smad3 and Smad4. Input lanes were loaded with 10% total protein compared to DNAP inputs. *C) EMSA of the* c-Jun *SBR.* Probes were combined with nuclear extract from TGF-β-induced and uninduced HaCaT cells, as indicated. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and peptide corresponding to the Mixer SIM were added as indicated. The positions of the TGF-β-induced Smad3-Smad4 complex and of a Smad4-containing complex that forms in the absence of TGF-β and is stabilised by an anti-Smad4 antibody are indicated.

to TGF-β (Bonni et al., 2001; Stroschein et al., 2001; Sun et al., 1999), activation of the pathway should lead to its removal form the *c-Jun* SBR and thus permit gene induction by active Smad complexes.

I tested the hypothesis that SnoN associates with the *c-Jun* SBR in the absence of TGF-β by DNAP of the *c-Jun* SBR with nuclear extract from HaCaT cells induced with TGF-β for one hour or left uninduced. This was followed by immunoblotting of crude extracts and pull-down eluates with an antibody specific to SnoN. Immunoblotting revealed that SnoN, as published in the literature, is degraded in response to TGF-β (Bonni et al., 2001; Stroschein et al., 2001; Sun et al., 1999). It also showed that SnoN associates with the *c-Jun* SBR in the absence of TGF-β and that this association is lost after one hour of signalling, probably because of SnoN degradation (Fig. 3.6A). It is likely that, as SnoN is re-synthesised at later time points, it is recruited again to the SBR and attenuates transcription. This could explain the transient induction of *c-Jun* that occurs upon TGF-β stimulation.

SnoN cannot bind DNA directly and must be recruited to the *c-Jun* SBR in the by another factor. This could be Smad4 or some another protein that has yet to be identified (Fig. 3.6B). This second possibility is supported by data that, in cells where Smad4 is knocked down, SnoN continues to associate with the *c-Jun* SBR in the absence of TGF-β (Ross, S.J. and Hill, C.S.; unpublished data). This seems to indicate that SnoN is recruited by other means and that Smad4 functions in a different way. An alternative explanation is that low levels of Smad4 are sufficient for SnoN recruitment but, since the knock-down of Smad4 was very efficient, this does not seem likely.

3.8 Mapping the minimal *c-Jun* SBR

Having investigated the protein composition of the TGF-β-inducible complex that forms on the *c-Jun* SBR, it was decided to investigate the DNA sequences required for complex formation. The 55 base-pair SBR used as full-length probe in all analyses to this point (Fig. 1.9A) had not been shown to be the minimal element

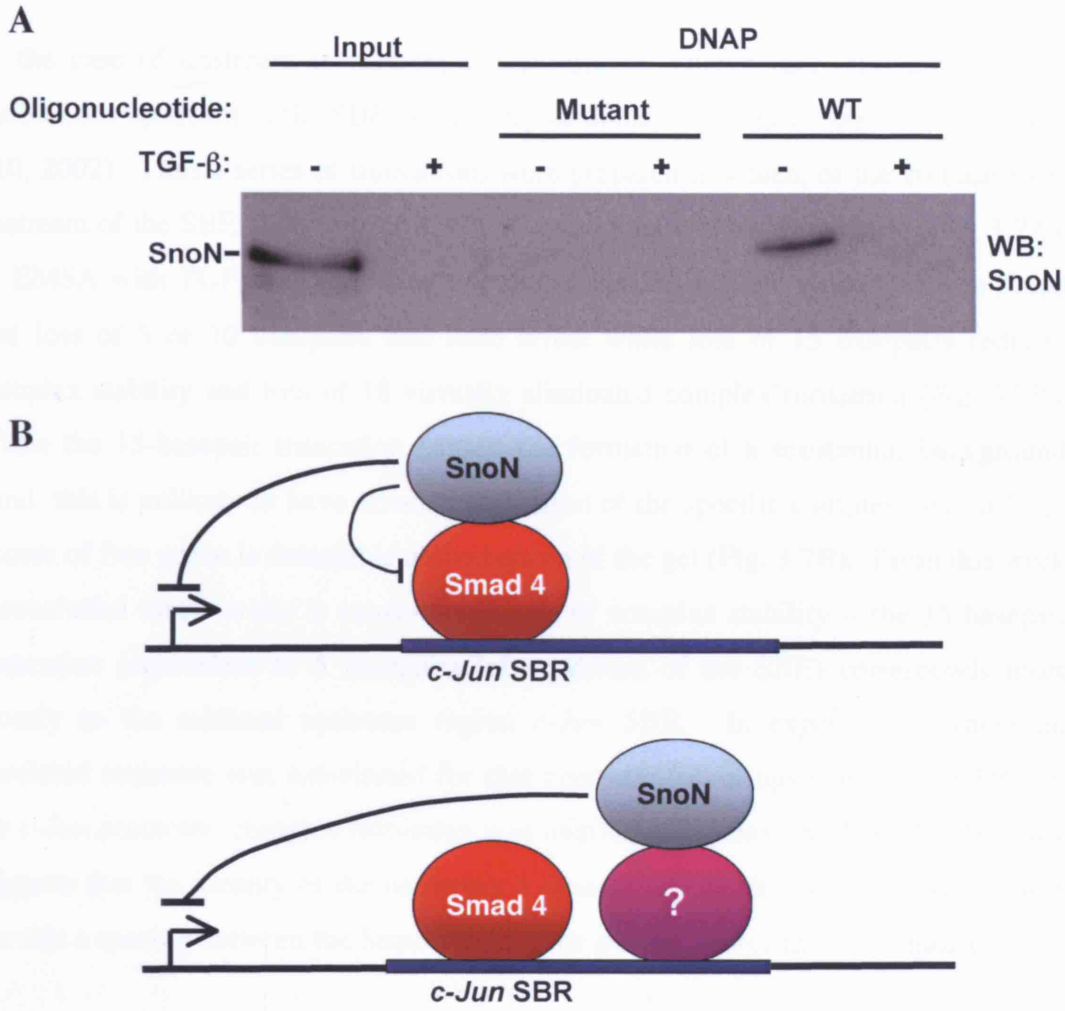


Fig. 3.6 SnoN associates with the *c-Jun* SBR in the absence of TGF-β. *A) DNAP of the* c-Jun *SBR and associated proteins.* Pull-downs were carried out as in Fig. 3.5 and precipitates were blotted with an antibody specific to SnoN. Input lanes were loaded with 10 % total protein compared to DNAP inputs. *B) Two alternative models of SnoN function on the* c-Jun *SBR.* SnoN is a Smad corepressor and, if it is recruited in this fashion on chromatin, probably functions to prevent activation of *c-Jun* transcription by Smad4 bound to the SBR in the absence of TGF-β. In one model, SnoN is recruited by Smad4 and represses *c-Jun* by recruiting the Sin3-HDAC complex and preventing recruitment of coactivators by Smad4. In an alternative model, another DNA-binding protein recruits SnoN, which in turn recruits the Sin3-HDAC complex to repress transcription. The well-documented ability of SnoN to interact directly with Smad4 supports the first model but there is experimental evidence in favour of the second model (see text). There is no evidence that SnoN might be able to bind DNA directly.

required for complex formation. Thus, I decided to map the minimal SBR. This was achieved by making a series of upstream and downstream truncations of increasing severity and using the resulting sequences as radiolabelled probes in EMSA.

In the case of unstream truncation, it was already known that elimination of all nucleotides upstream of the SBE completely abolishes complex formation (Inman and Hill, 2002). Thus a series of truncations were prepared in which, of the 20 base-pairs upstream of the SBE, deletions of 5, 10, 15 and 18 base-pairs were made (Fig. 3.7A). In EMSA with TGF-β-induced and uninduced HaCaT nuclear extract, it was found that loss of 5 or 10 basepairs had little effect while loss of 15 basepairs reduced complex stability and loss of 18 virtually eliminated complex formation (Fig. 3.7B). While the 15-basepair truncation caused the formation of a substantial background band, this is unlikely to have affected formation of the specific complex since a large excess of free probe is detectable at the bottom of the gel (Fig. 3.7B). From this work I concluded that – while it causes some loss of complex stability – the 15-basepair truncation (equivalent to 5 basepairs left upstream of the SBE) corresponds most closely to the minimal upstream region *c-Jun* SBR. In experiments where an unrelated sequence was substituted for that normally found upstream of the SBE in the *c-Jun* promoter, complex formation was unaffected (Inman and Hill, 2002). This suggests that the identity of the necessary 5 basepairs is not important, merely that it provide a spacing between the Smad binding site and the end of the DNA molecule.

The factor Y binding site in the *c-Jun* SBR was originally described as a CCAG triplet (Inman and Hill, 2002). However, there is a G immediately upstream of the first CCAG and it is possible to also view the sequence as a GCCA triplet. For reasons that will be explained below, this latter view has been adopted in this thesis. In the 55 base-pair SBR, there are 8 base-pairs downstream of the GCCA triplet. For my deletion analysis, radiolabelled probes were synthesised in which 2, 4, 6 or all 8 of these were deleted (Fig. 3.8A). When analysed by EMSA with TGF-β-induced and uninduced HaCaT nuclear extract I found that TGF-β-dependent complex formation could occur on all of these probes with some diminution on the probe with no base-pairs downstream of the GCCA triplet (Fig. 3.8B). I thus concluded that the 8-

A

FL: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

U1: 5'-TGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

U2: 5'-GGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

U3: 5'-CAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

U4: 5'-GCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

B

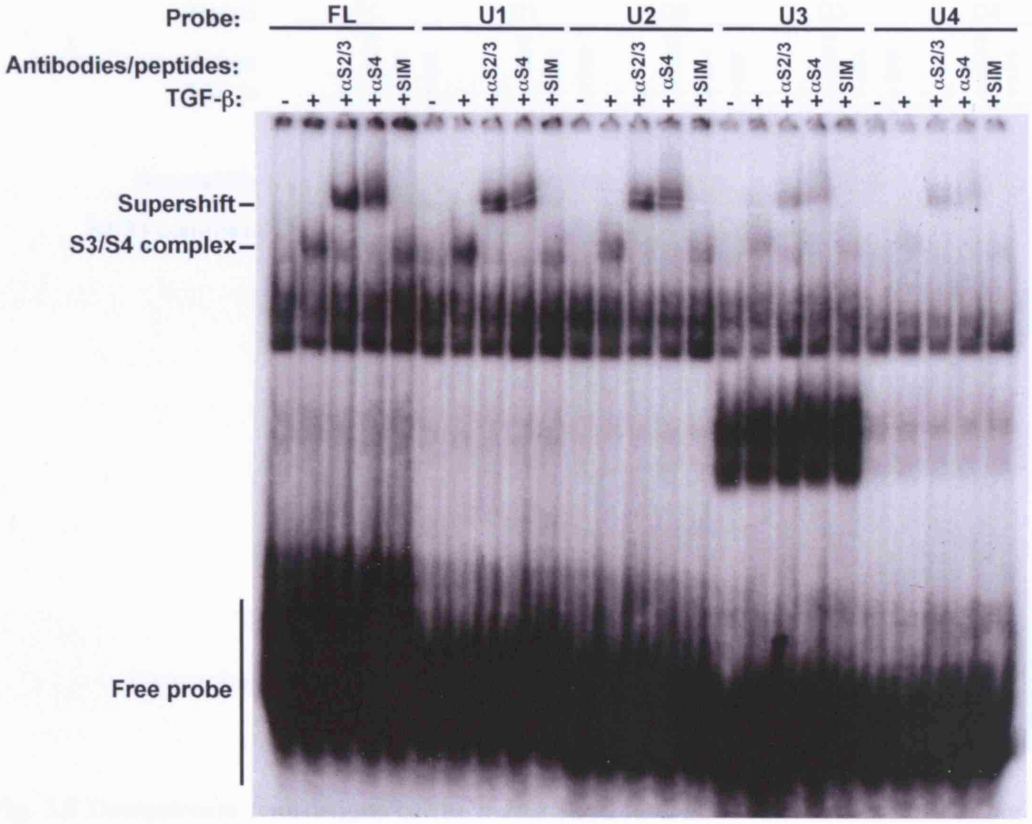


Fig. 3.7 Upstream truncations of the *c-Jun* SBR. *A) Upstream truncations of the 55 bp SBR used in previous studies.* These sequences were synthesised as radiolabelled EMSA probes. The original probe is designated full-length (FL) and truncations as U1-U4. The Smad-binding element (SBE) is indicated in blue and the factor Y binding site in red. *B) EMSA using the probes described in A.* EMSA was carried out using TGF-β-induced and uninduced nuclear extract from HaCaT cells, as indicated. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and a peptide corresponding to the Mixer SIM were added as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 (S3/S4) complex and of antibody-induced supershifts of this complex are indicated.

A

WT: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

D1: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGG-3'

D2: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTC-3'

D3: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGG-3'

D4: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCA-3'

B

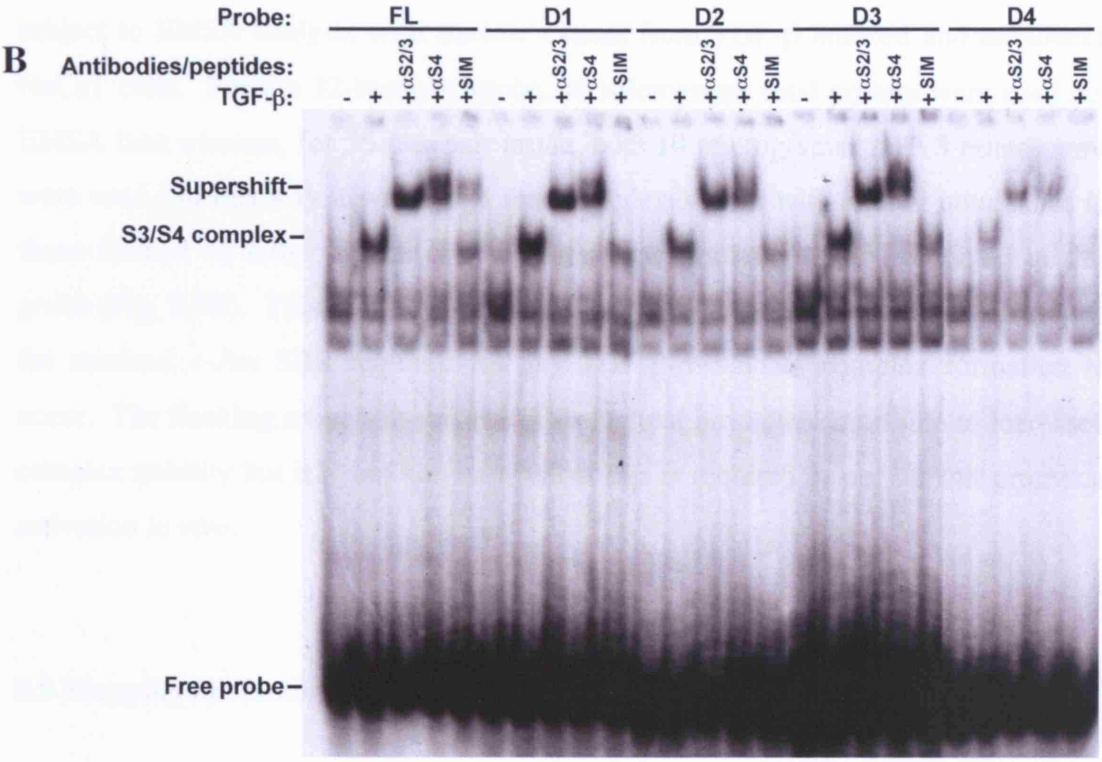


Fig. 3.8 Downstream truncations of the *c-Jun* SBR. *A) Downstream truncations of the 55 bp SBR used in previous studies.* These sequences were synthesised as radiolabelled EMSA probes. The original probe is designated full-length (FL) and truncations as D1-D4. The SBE is indicated in blue and the factor Y binding site in red. *B) EMSA using the probes described in A.* EMSA was carried out using TGF-β-induced and uninduced nuclear extract from HaCaT cells, as indicated. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and a peptide corresponding to the Mixer SIM were added as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 (S3/S4) complex and of antibody-induced supershifts of this complex are indicated.

basepair truncation corresponds most closely to the minimal downstream region of the *c-Jun* SBR.

The next issue to be considered was whether these two maximum truncations, when combined, have an additive effect on complex stability or whether they form a complex with similar abundance to the more destabilising of the two truncations. To test this, a 32-basepair probe was synthesised in which 5 basepairs were left unstream of the SBE and no sequences were left downstream of the GCCA triplet (Fig. 3.9A). This corresponds to a combination of the individual truncations at the upstream and downstream ends that appear to correspond to the minimal SBR. This was then subject to EMSA analysis with nuclear extract from TGF-β-induced and uninduced HaCaT cells. For the 32-basepair probe, 10 micrograms total protein were used per EMSA lane whereas, for 55-basepair probe, both 10 micrograms and 3.3 micrograms were used. In this way it was found that complexes form with similar abundance to those formed on either of the two separate truncations that were combined in this probe (Fig. 3.9B). This indicates that this 32-basepair element corresponds closely to the minimal *c-Jun* SBR required for any TGF-β-dependent complex formation to occur. The flanking sequences present in the natural promoter contribute to increased complex stability but it is not known whether this is required or not for full promoter activation *in vivo*.

3.9 Mapping the factor Y binding site

Knowing the sequences that comprise the minimal *c-Jun* SBR, it next became important to precisely map the factor Y binding site. This was of great importance since a deeper understanding of the sequence of this site could provide a clue as to the identity of factor Y.

From the work of Inman and Hill, it was already known that no complex forms on a probe in which the terminal CA sequence of all three GCCA repeats is changed to an AT (Inman and Hill, 2002). However, this mutational analysis did not reveal the true extent of the factor Y binding site. It remained possible, for example that only one or

A

FL: 5'-GGAGGTGCGCGGAGTCAGGCAGACAGACAGACACAGCCAGCCAGCCAGGTCGGCA-3'

Min: 5'-CAGGCAGACAGACAGACACAGCCAGCCAGCCA-3'

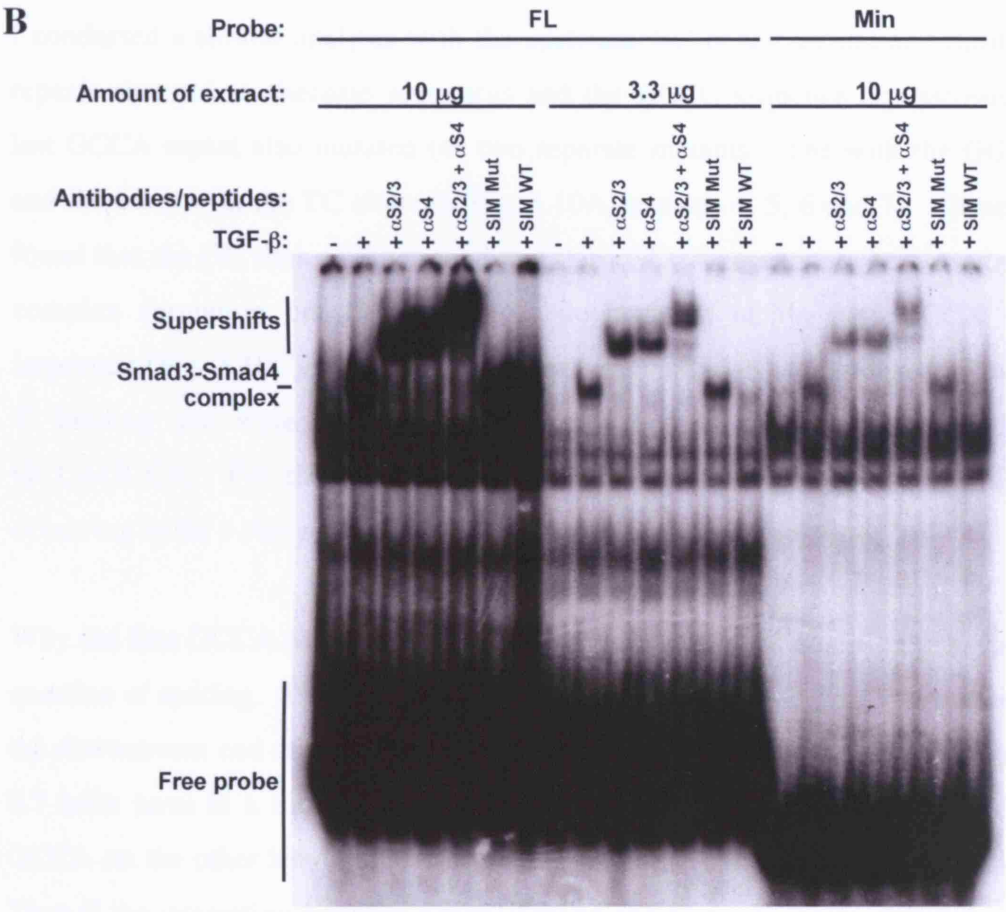


Fig. 3.9 The minimal *c-Jun* SBR. *A) "Full-length" (FL) and minimal (Min)* c-Jun *Smad-binding regions.* These sequences were synthesised as radiolabelled EMSA probes. The SBE is indicated in blue and the factor Y binding site in red. *B) EMSA using the probes described in A.* EMSA was carried out using TGF-β-induced and uninduced nuclear extract from HaCaT cells, as indicated.. For FL probe, 10 μg and 3.3 μg total protein were used. For Min probe, 10 μg was used. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and wild-type (WT) and mutant (Mut) peptides corresponding to the Mixer SIM were added as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated. Levels of complex formed on the minimal probe are estimated to be approximately one-sixth those formed on the full-length probe.

two of the three two-base changes made in the sequence were important. As a consequence, I decided to map the precise sequences required for factor Y binding. My initial investigation involved separately changing the terminal CA of each of the three GCCA repeats to AT (Fig. 3.10A, mutants 1, 2 and3) and analysing the resulting probes by EMSA. In these experiments, it was found that the second and third CA sites contribute to complex stability but the first one does not (Fig. 3.10B).

I conducted a similar analysis with the upstream GC of the second and third GCCA repeats changed to alternate sequences and the GGTC sequence downstream of the last GCCA repeat also mutated (as two separate mutants – one with the GG altered and the other with the TC altered) (Fig. 3.10A, mutants 4, 5, 6 and 7). These studies found that the GC sites of the second and third GCCA repeats are both important for complex formation but that sequences downstream of the last GCCA are not important (Fig. 3.10C). In this way the upstream and downstream ends of the factor Y binding site were mapped and the sequence of the site was found to be GCCAGCCA. For this reason it was decided to view the factor Y binding site occurring in the *c-Jun* promoter as consisting of two GCCA repeats.

Why the first GCCA repeat is not important is unclear but it could, in principle, be a question of spacing. The upstream end of the second GCCA repeat is separated from the downstream end of the SBE by 7 basepairs, which is equivalent to approximately 0.7 helix turns in a canonical B-DNA double helix. The upstream end of the first GCCA on the other hand is located only 3 basepairs (0.3 helix turns) from the SBE. Thus if the interaction between Smad3/Smad4 and factor Y required that their DNA binding domains lie on opposite sides of the DNA duplex then the first GCCA repeat would be positioned incorrectly for the interaction to take place.

This view is, however, contradicted by evidence that any two of the three AGAC elements in the *c-Jun* SBE are sufficient for full TGF-β-dependent complex formation on the SBR (Inman and Hill, 2002). In situations where only the first two AGACs are intact, the spacing between the functional SBE and the start of the second GCCA repeat is completely different to situations where only the second two AGACs are available. It could be argued that different GCCA repeats are utilised when different

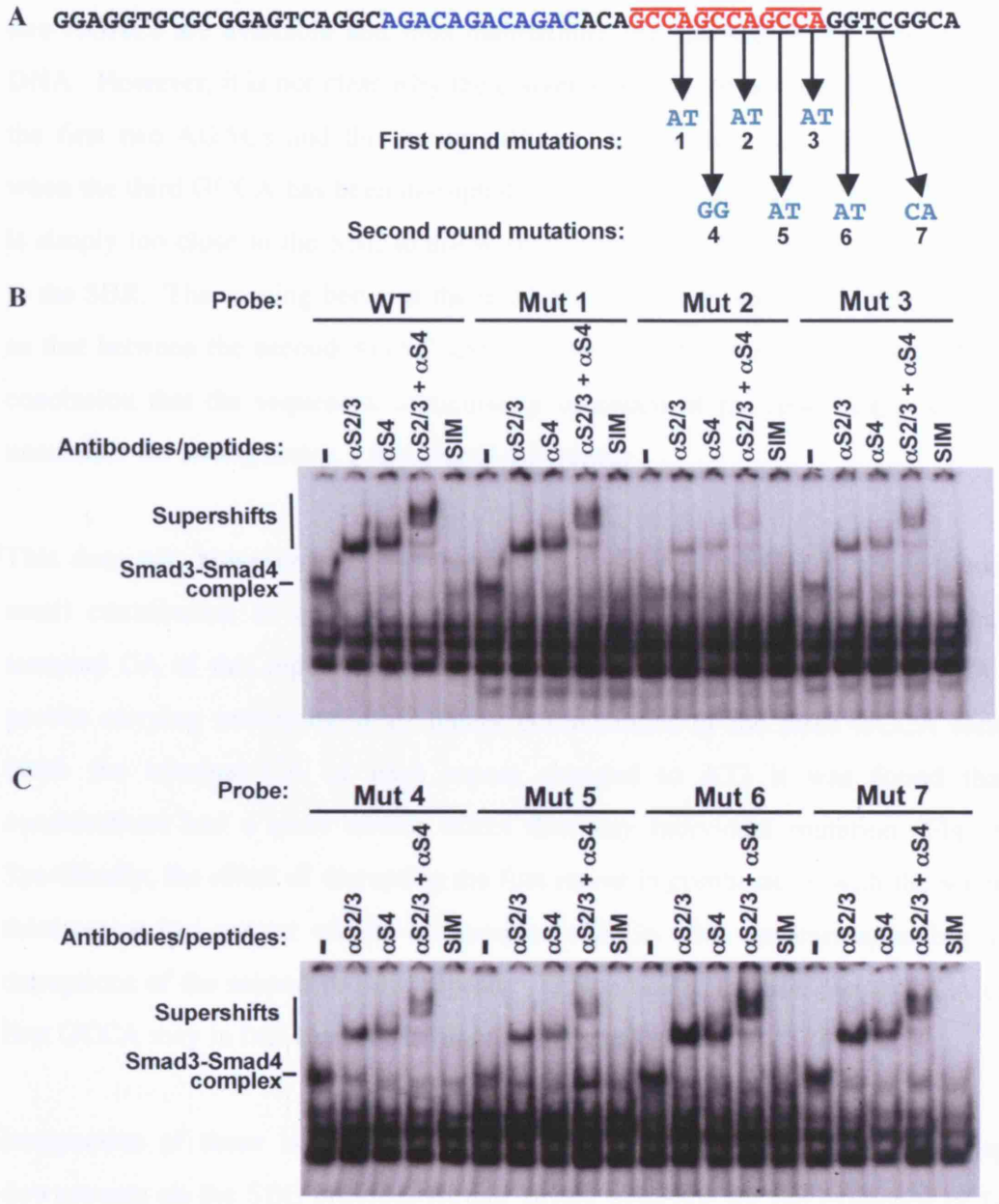


Fig. 3.10 Factor Y requires the second and third repeats of the $(GCCA)_3$ sequence in the *c-Jun* SBR. *A) Mutagenesis of the c-Jun SBR.* The analysis was carried out in two rounds with the sequence being substituted as indicated. All sequences were synthesised as radiolabelled probes for EMSA. In the case of mutant 4, it was important that a substitution to AT not be used since this was found to create a binding site for an unknown protein that efficiently titrates the probe (data not shown). Substitution to GG was used instead. *B) and C) EMSA carried out using the probes described in A.* Probes were combined with nuclear extract from TGF-β-induced HaCaT cells. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and a peptide corresponding to the Mixer SIM were added as indicated. The positions of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated. The data show that complex formation is reduced on mutants 2, 3, 4 and 5, indicating that the factor Y binding site corresponds to the second and third GCCA repeats.

AGACs are available – factor Y, say, binding to the first GCCA when only the first two AGACs are available and thus maintaining the spacing of 0.7 helical turns of DNA. However, it is not clear why the converse would not be true: Smads binding to the first two AGACs and this being sufficient to maintain full complex formation when the third GCCA has been disrupted. It also seems unlikely that the first GCCA is simply too close to the SBE to allow simultaneous binding of Smads and factor Y to the SBR. The spacing between the third AGAC and the second GCCA is the same as that between the second AGAC and the first GCCA. We are thus brought to the conclusion that the sequences immediately upstream of the first GCCA are simply unsuitable for strong factor Y binding to this repeat.

This does not, however, rule out the possibility of the first GCCA repeat making a small contribution to complex stability. In fact, I uncovered evidence that the terminal CA of this repeat does indeed make a slight contribution. In EMSA with probes carrying mutations in all binary combinations of the three GCCA elements (with the terminal CA of each repeat changed to AT) it was found that all combinations had a more severe effect than any individual mutation (Fig. 3.11). Specifically, the effect of disrupting the first repeat in combination with the second or third repeat had greater effects on complex stability than the corresponding single disruptions of the second or third repeats. This indicates that the terminal CA of the first GCCA may in fact have some functional significance.

Irrespective of these issues, the fact that a clear binding site has been mapped downstream on the SBE provides further strong evidence for the existence of factor Y. It does not however immediately facilitate identification of this factor since none of the ways in which the site can be viewed on the basis of current data – GCCAGCCA, CCAGCC or (CA)GCCAGCCA – bear close similarity to any known transcription factor binding site. Before the site was mapped, an overlapping site with 50% similarity to a consensus nuclear factor 1 (NF1) binding site was identified by computational analysis of the sequence. However, DNAP and EMSA assays for NF1 binding gave negative results and the putative site does not correspond to the results of careful mapping (data not shown).

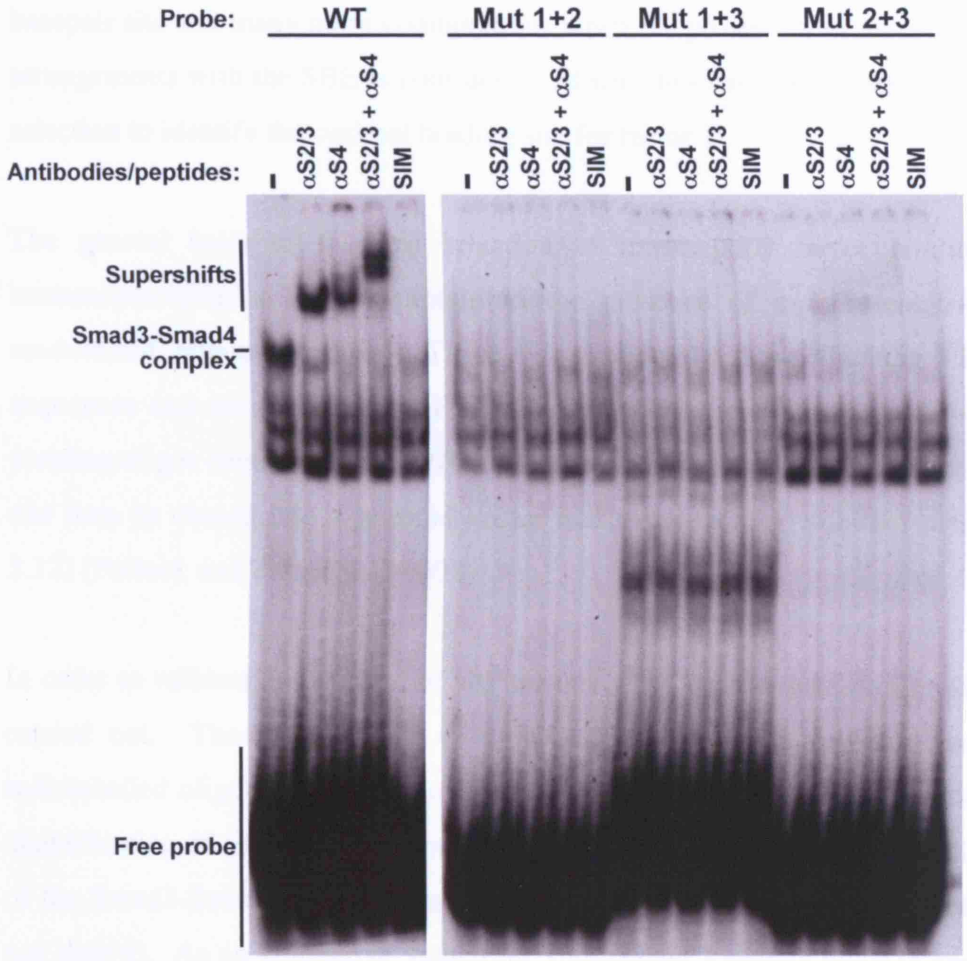


Fig. 3.11 Disruption of any two GCCA repeats in the *c-Jun* SBR completely disrupts complex formation. EMSA was carried out with nuclear extract from TGF-β-induced HaCaT cells and probes with indicated combinations of the mutations present in the probes shown in Fig. 3.10A. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and a peptide corresponding to the Mixer SIM were added as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated.

3.10 An attempt to determine the physically optimal binding site for factor Y

Mapping of the factor Y binding site revealed an 8- or possibly 10-basepair site. Biologically, the sequence of this depends on a GCCA double repeat but this may not be the physically optimal binding site and knowledge of such a site could facilitate identification of factor Y. There are 4^8 (or 65,536) possible sequences for an 8-basepair site and many more combinations when the possibility of alternative spacing arrangements with the SBE is considered. It was thus decided to attempt *in vitro* site selection to identify the optimal binding site for factor Y.

The general basis of *in vitro* selection of transcription factor binding sites is immunoprecipitation of the protein in the presence of a representative pool of randomised oligonucleotides. These oligonucleotides have flanking PRC primer sequences and are amplified by PCR after IP. This cycle is repeated several times, yielding oligos containing high affinity binding sites for the protein of interest. These can then be cloned into a plasmid vector and sequenced by standard methods (Fig. 3.12) (Pollock and Treisman, 1990).

In order to validate the concept of site selection by this method, McKay assays were carried out. These involve IP of a DNA-binding protein in the presence of a radiolabelled oligonucleotide corresponding to its cognate binding site, followed by quantification of co-precipitated probe (Pollock and Treisman, 1990). McKay assays of the Smad3-Smad4 complex that forms on the *c-Jun* SBR were unsuccessful (data not shown). As an alternative validation, flag-tagged *Xenopus* FoxH1a and FoxH1b were synthesised by *in vitro* translation in rabbit reticulocyte lysate and were each immunoprecipitated in the presence of three alternative radiolabelled oligonucleotides. One of these corresponded to the wild-type ARE element and the other two to the products of site selection with each of FoxH1a and FoxH1b (Howell, M. and Hill, C.S.; unpublished data).

McKay assays carried out under these conditions were washed and eluted by boiling the washed beads in SDS sample buffer. Eluates were then analysed by immunoblotting with an antibody specific to the flag tag and by Cerenkov counting of

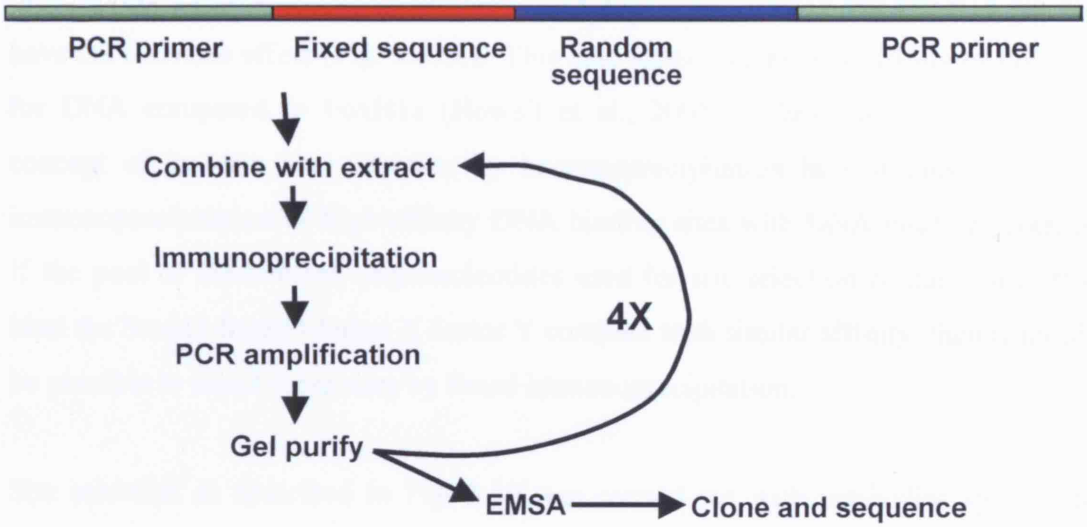


Fig. 3.12 ***In vitro*** **selection of the physically optimal DNA sequences for binding of a specific protein.** A representative pool of random-sequence oligonucleotides with flanking PCR primers is combined with extract and immunoprecipitation of the relevant protein is carried out. The DNA co-precipitated in this way is amplified by PCR and purified. This cycle is repeated several times until a probe that forms specific complexes with the protein of interest in EMSA is obtained. This probe can then be cloned and sequenced to reveal the physically optimal binding site for the factor under investigation. The oligonucleotide used may contain a fixed binding site adjacent to a randomised region or the sequence between the primers may be entirely random.

the radiolabelled probe. This showed that *in vitro* translated FoxH1 proteins were pulled-down efficiently in all conditions (Fig. 3.13B) and that radiobalelled DNA was co-precipitated in a very specific manner (Fig 3.13A). The ARE element was not co-precipitated with either protein but each protein co-precipitated the sequence that had been generated by earlier site selections with that same protein. FoxH1a co-precipitated the product of a site-selection with FoxH1a and FoxH1b the product of a site selection with FoxH1b. Interestingly, FoxH1b also precipitated a smaller amount of the probe originally derived from site selection with FoxH1a but FoxH1a did not have the converse effect (Fig. 3.13A). This may reflect the greater affinity of FoxH1b for DNA compared to FoxH1a (Howell et al., 2002). These assays validate the concept of *in vitro* site selection by immunoprecipitation in that they show co-immunoprecipitation of high-affinity DNA binding sites with DNA binding proteins. If the pool of randomised oligonucleotides used for site selection contains sites that bind the Smad3-Smad4-factor X-factor Y complex with similar affinity, then it should be possible to select these sites by Smad immunoprecipitation.

Site selection as described in Fig. 3.12 was carried out with antibodies specific to Smad2/Smad3 or Smad3 only and a radiolabelled probe. This probe contained the AGAC triplet and 10 basepairs of sequence that lie immediately upstream of it in the *c-Jun* SBR. The 20 basepairs downstream of the AGAC triplet were randomised and this entire sequence was positioned between two 25-basepair PCR primer sequences. Immunoprecipitation was carried out in the presence of nuclear extract from TGF-β-induced and uninduced HaCaT cells, followed by removal and washing of the beads and elution with detergent. The co-precipitated DNA was then isolated by polyacrylamide gel electrophoresis and subjected to PCR amplification. As with co-precipitated DNA, PCR products were isolated on polyacrylamide gels, enabling them to be used in the next round of selection.

This procedure was repeated for four cycles and the final products were analysed by EMSA. On probes selected in the presence of TGF-β-induced extracts, no formation of TGF-β-dependent complexes was detected in this way. Instead a number of high-abundance TGF-β-independent complexes were formed on each probe (Fig. 3.14). Site selections that have been successful in other studies have generally involved

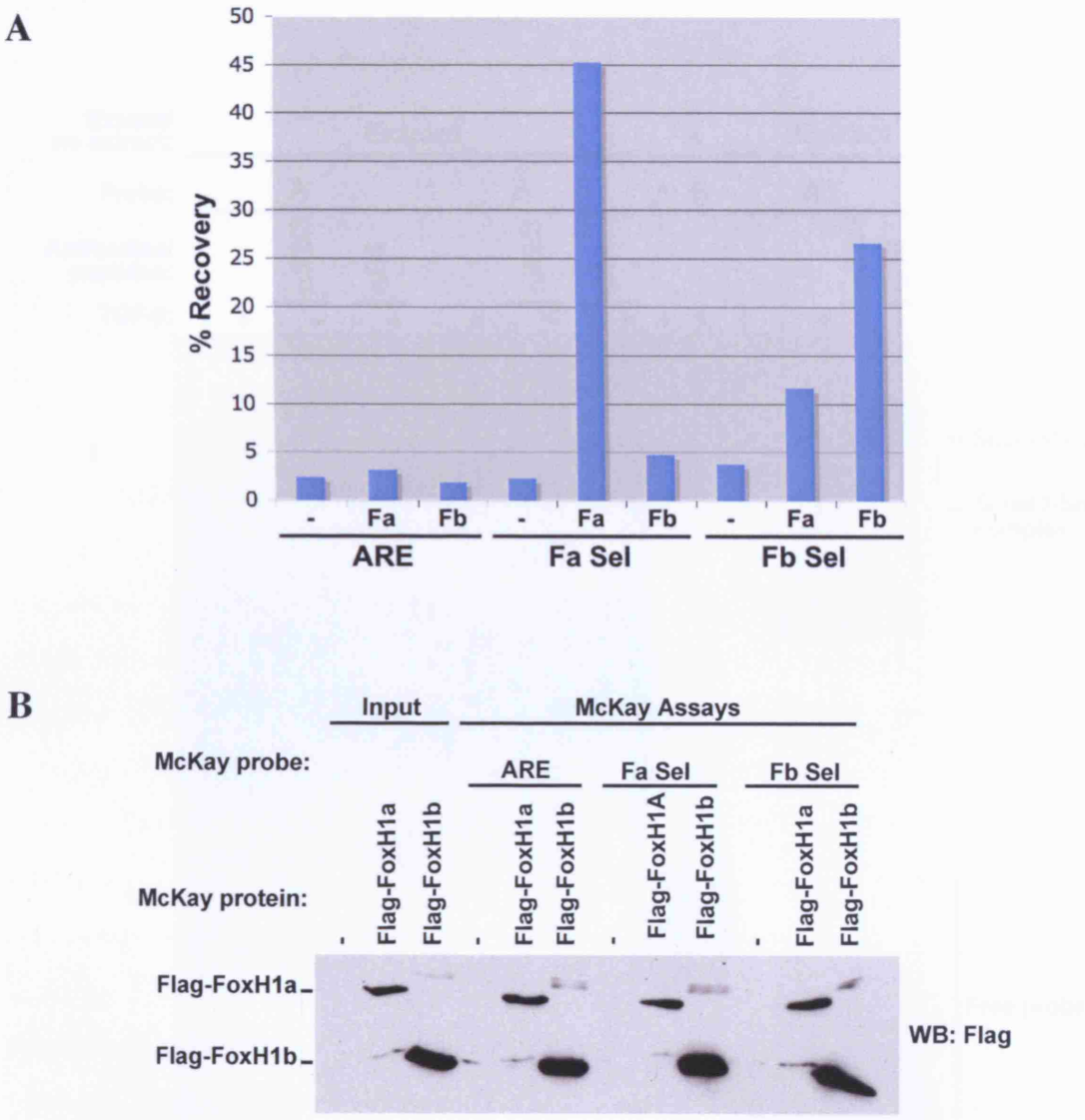


Fig. 3.13 Ratification of the site selection protocol: immunoprecipitation of DNA-binding proteins co-precipitates very specific sequences. *A) IP of Flag-FoxH1a co-precipitates optimised FoxH1a sites and IP of Flag-FoxH1b co-precipitates optimised sites for both FoxH1a and FoxH1b.* McKay assays involve IP of DNA-binding proteins in the presence of a radiolabelled probe corresponding to the DNA binding site of that factor. In this case McKay assays were carried out with *in vitro* translated Flag-FoxH1a and Flag-FoxH1b combined with radiolabelled probes corresponding to the activin response element (ARE) or probes derived from a site selection with FoxH1a (Fa Sel) or FoxH1b (Fb Sel). McKay assays were eluted with flag peptide and eluates were subjected to Cerenkov counting. *B) Immunoblot of McKay eluates and inputs with an antibody specific to the Flag tag.*

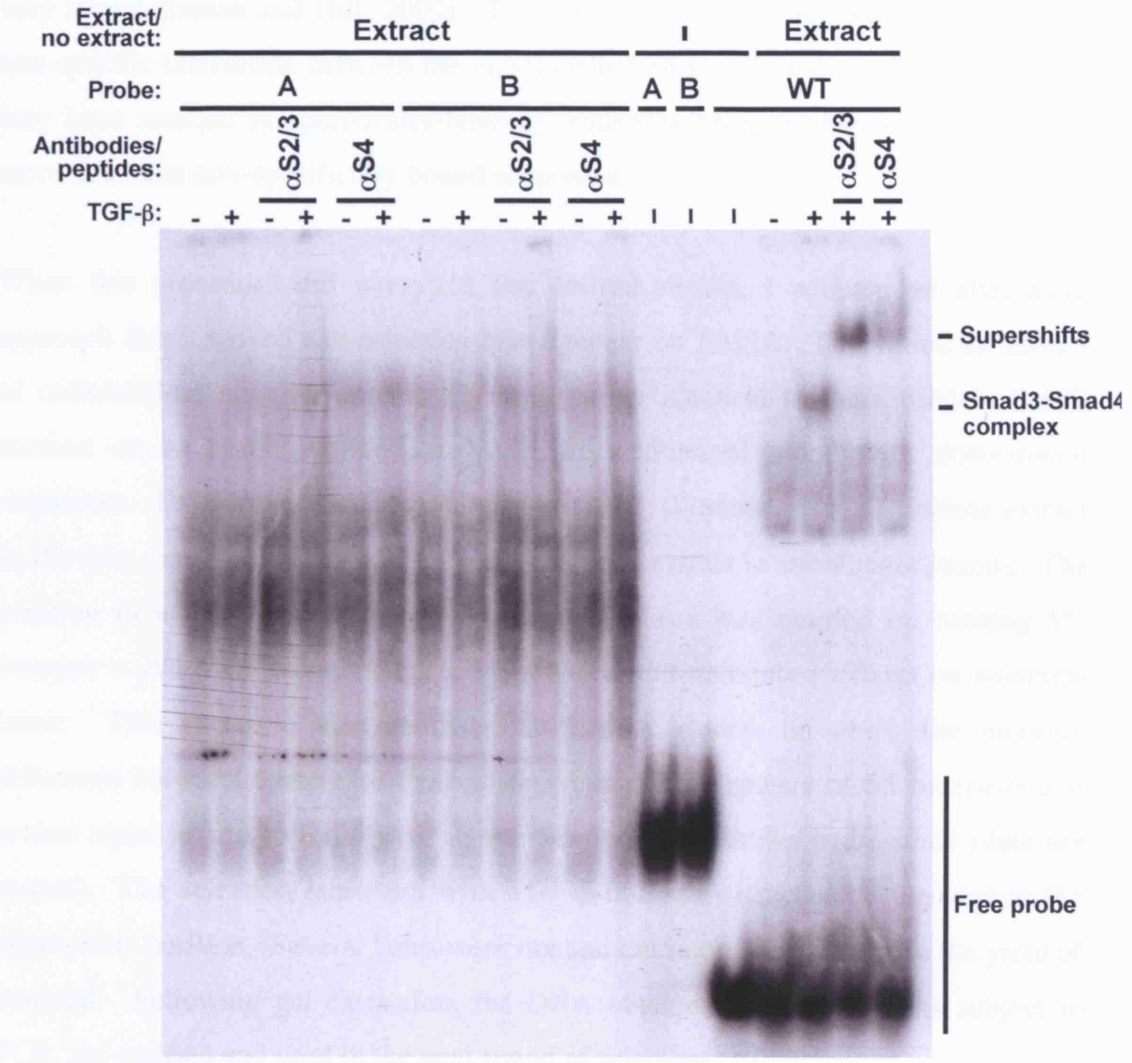


Fig. 3.14 Site selection as carried out as in Fig. 3.13 does not isolate sequences upon which TGF-β-dependent complexes can form. Site selection was carried out with TGF-β -induced extract and with antibodies specific to Smad3 (condition A) and Smad2/Smad3 (condition B). Probes derived from these selections were used in EMSA alongside wild-type (WT) probe. TGF-β-induced and uninduced extracts were used in this analysis and antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) were added as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated. Probes A, B and WT were also run in the absence of extract, as indicated.

immunoprecipitation of transcription factors that bind DNA strongly in the absence of associated factors (Pollock and Treisman, 1990) and have often used overexpression of these factors or addition of recombinant proteins. In the present study, a more elaborate complex was being pulled down and its affinity for DNA is known not to be very strong (Inman and Hill, 2002). Together with the fact that there was clearly a non-specific interaction between the radiolabelled oligonucleotide and the beads, this may have resulted in specifically-binding sequences being unable to compete with more abundant non-specifically bound sequences.

When this procedure did not yield the desired results, I adopted an alternative approach that involved site-selection based purely on EMSA. This relied on EMSA of radiolabelled oligonucleotides that were either identical to those used in the IP method or in which the SBE region was randomised along with downstream sequences. EMSA was carried out, with just TGF-β-induced HaCaT nuclear extract in the first round and with induced and uninduced extract in subsequent rounds. The position in which the complex was predicted to run was marked by running 55-basepair *c-Jun* SBR probe with TGF-β-induced and uninduced extract on adjactent lanes. This approach was validated by EMSA studies in which the mobility difference between complexes formed on *c-Jun* SBR elements of 55 basepairs and probes equal in size to those used in site selection was shown to be small (data not shown). The selection lanes run with TGF-β-induced extract were then cut at the appropriate position. Several lanes were run and cut in order to maximise the yield of material. Following gel extraction, the DNA obtained in this way was subject to PCR, gel purified and used in the next round of selection.

One of the advantages of this approach is that, because the EMSA gels must be visualised by autoradiography before gel slices are cut, a read-out of the success or failure of the procedure automatically built into the system. However, it was observed that, even after three rounds of selection, no TGF-β-dependent complexes had emerged. Instead, a range of abundant TGF-β-independent complexes were evident in all rounds of selection and diminished in number as the procedure continued. Thus, the failure of the procedure was probably the result of low

abundance of the target complex in the first round and an inability to compete efficiently with much stronger complexes.

3.11 Conclusions

The work described in this chapter deepens our understanding of TGF-β-dependent complex formation on the *c-Jun* promoter. It confirms the presence of factors X and Y and provides a more detailed understanding of the DNA sequences required for complex formation. Specifically it shows that the minimal SBR is 32 basepairs in length and that this encompasses an AGAC triplet to which Smad3 and Smad4 bind and a GCCA double repeat downstream of this to which factor Y binds. It also shows that Smad4 and SnoN, but not Smad3 associate with the *c-Jun* SBR in the absence of TGF-β signalling. The possible significance of these findings is discussed in Chapter 5 below.

There are a number of further studies that could be carried out to further characterise TGF-β-dependent complex formation on the *c-Jun* SBR. These include further validation of the existence of factor X using EMSA and McKay assays with wild-type Smad3 and Smad3 mutated in the SIM-binding pocket. However, little more is likely to be discovered about this complex without identification of factors X and Y and any other proteins that may be present. The development of technology to achieve this goal is the main area of work described in the next chapter.

PURIFICATION STRATEGY TO ISOLATE THE TGF-β-DEPENDENT COMPLEX THAT BINDS THE *c-Jun* SBR

4.1 Introduction

Despite a large number of reports concerning specific interactions between Smad molecules and regulatory transcription factors, identification of such factors remains a major challenge (see Chapter 1 above). Only a few of the reported partners have been properly validated and very few interactions have been well characterised on the molecular level. The work described in the preceding chapter confirms that there are additional factors present in the TGF-β-dependent complex that forms on the *c-Jun* SBR. One of these – factor X – apparently contains a motif functionally equivalent to the Mixer SIM, a feature not previously identified in any factor expressed in the mammalian adult. The other unidentified component – factor Y – binds a GCCA double repeat that bears little similarity to any previously identified transcription factor binding site. It is thus likely that these factors represent novel Smad partners and it was decided that they should be purified, identified and characterised.

4.2 Protein-DNA crosslinking to facilitate identification of complex components

Before embarking on purification, I decided to ascertain the number and molecular masses of the DNA-binding proteins present in the complex. As well as providing yet more evidence for the existence of non-Smad subunits, this would provide a useful guide to the analysis of purification eluates.

The method adopted to obtain this information was protein-DNA crosslinking by UV irradiation (Chodosh et al., 1986). This involves synthesis of a radiolabelled DNA probe containing the binding site of interest but with bromodeoxyurindine (BrdU) substituted in place of thymidine. Noncovalent protein-DNA complexes are allowed to form on the probe before irradiation with UV light. This causes BrdU free radicals to form in the probe. These then form photoadducts with various amino acids in

DNA-bound proteins. Such crosslinks can be formed only with residues that already make direct contact with BrdU and are formed at such low efficiency that no more than one protein-DNA crosslink is formed per noncovalent complex. Once UV irradiation is terminated, complex components are analysed by SDS-PAGE followed by autoradiography. This detects adducts of DNA-binding proteins with radiolabelled single strands of DNA and allows the molecular mass of unknown DNA-binding proteins to be determined.

This procedure was attempted for the TGF-β-dependent complex that forms on the *c-Jun* SBR. Probes were synthesised as described above and EMSA was carried out with nuclear extract from TGF-β-induced and uninduced HaCaT cells. A single supershift was also carried out with an antibody specific to Smad2/Smad3. Gels were then UV irradiated in a uniform fashion, dried and visualised by autoradiography. This revealed formation of the expected TGF-β-dependent complexes containing Smad3 (data not shown). Supershifted and non-supershifted complexes were subsequently excised and gel slices were placed into the wells of an SDS-PAGE gel. This was then run, dried and visualised by autoradiography.

Unfortunately, no specific bands were detectable in this second-dimension analysis. Instead, autoradiograms revealed a uniform smear from top to bottom of all lanes (data not shown). The reason for this failure is unclear but it is possible that, even at the low energies and durations used, UV irradiation was too efficient. A diverse range of multiply crosslinked products may have been formed. These could have involved links between DNA strands, between DNA and multiple protein molecules and even between DNA and some component of the gel itself, such as glycerol or unpolymerised acrylamide. This conclusion is supported by experiments in which nuclear extract was omitted and the probe alone was subject to EMSA, crosslinked, excised and analysed by SDS-PAGE. This revealed a smear that is more concentrated at the bottom of the gel (data not shown). In spite of this, reduction of irradiation time to 10 minutes did not improve the performance of protein-DNA crosslinking. Given these results, I decided not to seek any further information regarding the components of the complex and to press ahead with development of a purification system for their direct identification.

4.3 Single-step DNA pull-down is not a viable purification strategy

The most obvious way to isolate components of the complex of interest was DNA pull-down using the *c-Jun* SBR as bait. Given the fact that the complex is stable only at low salt concentrations (Inman, GJ and Hill, CS; unpublished data) and the fact that higher salt concentrations are often required to prevent non-specific binding of background proteins to purification matrices, I realised that high background could be a serious problem in this approach. Nevertheless, because it would represent a simple and rapid purification strategy, I decided to explore the option of preparative single-step DNA pull-down (DNAP).

This was carried out in a similar fashion to the analytical DNA pull-downs already described. Double-stranded oligonucleotide corresponding to the *c-Jun* SBR was synthesised with one strand specifically biotinylated at the 5' end. This was immobilised on neutravidin beads and the beads were washed to remove unbound oligonucleotide. DNA-coated beads generated in this way were then incubated with nuclear extract from TGF-β-induced or uninduced HaCaT cells. After incubation, beads were removed, washed and eluted by incubation with 600 mM NaCl.

Eluates obtained in this way were then analysed by EMSA and by SDS-PAGE with silver staining of resolved proteins. EMSA revealed the presence of the target complex in the eluate, indicating that all factors essential for its formation are retained in DNAP (Fig. 4.1A). SDS-PAGE, however, showed that a very large number of other proteins – probably numbering several hundred – are also retained (Fig. 4.1B). These are pulled-down independently of TGF-β and thus represent nonspecifc background proteins, rather than *bona fide* components of the complex. The background problem was not significantly resolved by pre-blocking the beads with bulk proteins such as casein or lysozyme (data not shown). Neither was it solved by pre-clearing extracts on neutravidin beads lacking immobilised DNA (data not shown). Raising the NaCl concentration was not tested as a solution to the background problem since it was known that the necessary increases would completely disrupt the target complex (Inman, GJ and Hill, CS; unpublished data).

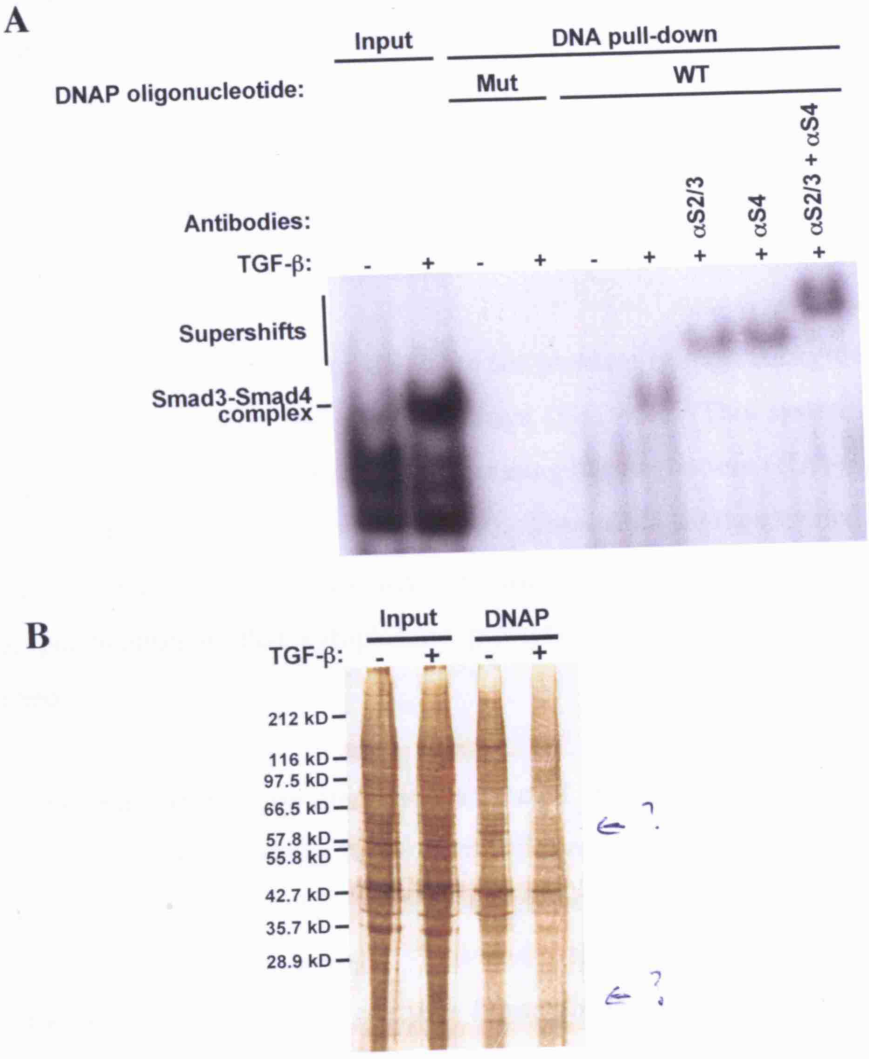


Fig. 4.1 DNA pull-down using the *c-Jun* SBR as bait retains all factors required for TGF-β-dependent complex formation on this element along with many other proteins. *A) Retention of essential complex components.* DNAP was carried out as described in Fig. 3.5A using wild-type (WT) and mutant (Mut) *c-Jun* SBR oligonucleotide and nuclear extract from TGF-β-induced and uninduced HaCaT cells. DNAPs were eluted with buffer containing 600 mM NaCl, diluted and analysed by EMSA using a radiolabelled probe corresponding to the *c-Jun* SBR. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) were included as indicated. The positions the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated. *B) Retention of other proteins.* DNAP was carried out as described in Fig. 3.5A using wild-type bait corresponding to the *c-Jun* SBR and nuclear extract from TGF-β-induced and uninduced HaCaT cells (2 mg total protein). Elution was by incubation with 600 mM NaCl. Eluates were concentrated by ultrafiltration and analysed by SDS-PAGE. Gels were then silver stained to detect retained proteins. Input lanes were loaded with 10% total protein compared to DNAP inputs

I thus concluded that, with a complex that is only stable at low salt concentrations, the background problem precludes adequate purification in a single step. I decided to develop a more advanced purification system incorporating multiple steps.

4.4 Two-step affinity purifiction of the TGF-β-dependent complex that forms on the *c-Jun* SBR

The system that was designed to eliminate the problem of high background involved two affinity steps carried out in batch format (Fig. 4.2). This system requires the generation of a HaCaT cell line stably expressing haemagluttinin (HA)-tagged Smad3 at endogenous levels (see section 4.5 below). These cells are then treated with TGF-β or left untreated and nuclear extract is prepared. Nuclear extraction gives a crude initial purification in that cytoplasmic proteins are depleted and nuclear proteins enriched.

These nuclear extracts are immunoprecipitated in the presence of a biotinylated oligonucleotide corresponding to the *c-Jun* SBR with and antibody specific to the HA tag (Fig. 3.1A), washed and eluted by multiple sequential incubations with a peptide corresponding to the HA epitope. This is the first affinity step. Inclusion of the oligonucleotide is essential for complex formation and facilitates the second affinity step (see below). Competitive elution with HA peptide is also important since it allows all components of the complex (including DNA) to be released from the beads under non-denaturing conditions but does not facilitate release of non-specificlly bound background proteins. In addition, anti-HA antibodies are cheap and readily available and can thus be used in large quantities.

The second step involves DNA pull-down of the biotinylated oligonucleotide as described in Fig. 3.5A. This is facilitated by retention of this oligonucleotide after the first step and simply requires addition of neutravidin beads. After incubation of this mixture, beads are removed, washed at moderate salt levels and eluted with 600 mM NaCl. Final eluates are then analysed by immunoblotting and EMSA to confirm the presence of HA-Smad3, Smad4, factor X and all other essential complex

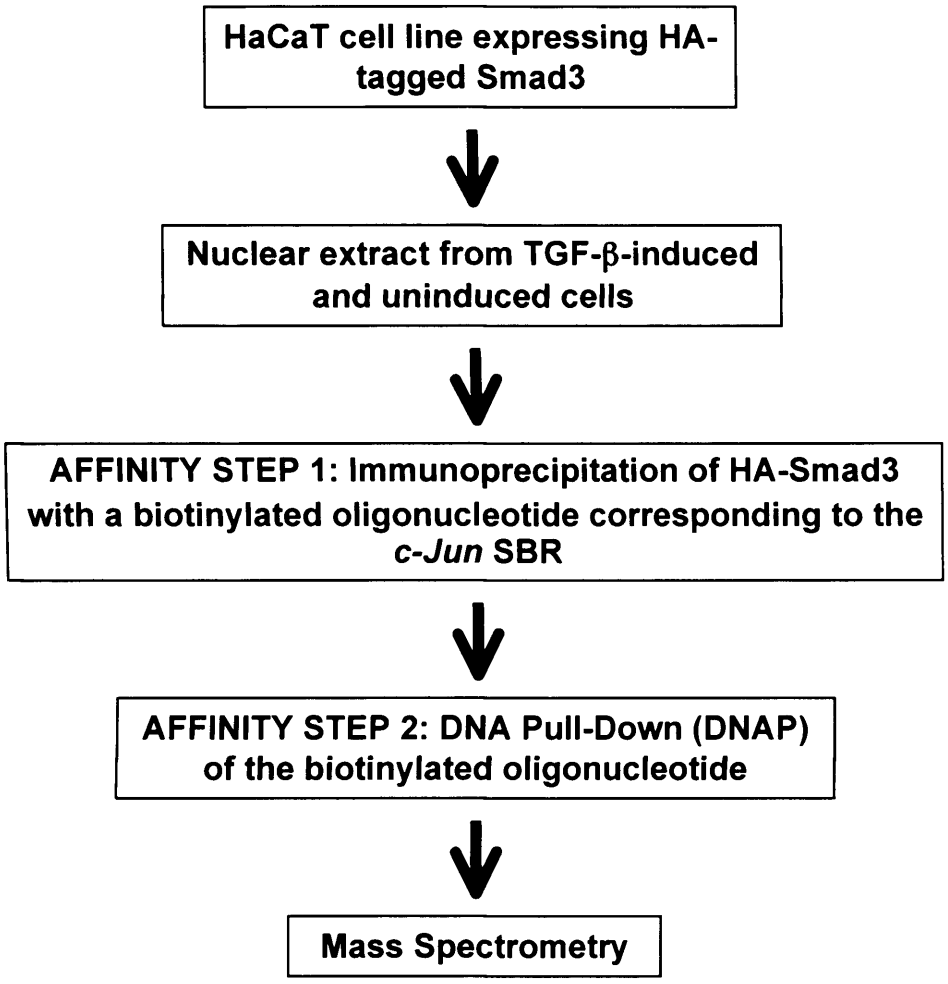


Fig. 4.2 Two-step affinity purification strategy designed to isolate all factors associated with the *c-Jun* SBR in the presence of TGF-β. A HaCaT cell line stably expressing haemaglutinin (HA)-tagged Smad3 is treated with TGF-β or left untreated. Nuclear extract is prepared and subjected to anti-HA IP in the presence of wild-type and mutant biotinylated oligonucleotides corresponding to the *c-Jun* SBR (see Fig. 3.1A). The IP is washed and eluted by multiple sequential incubations with a peptide corresponding to the HA epitope. The eluates thus obtained are then used as the input for the second affinity step: DNAP using neutravidin-coated beads to capture the biotinylated oligonucleotide. The DNAP is then washed and eluted with buffer containing 600 mM NaCl. The final eluate is then analysed by mass spectrometry to identify the proteins present. Proteins present in the final eluate under all conditions (± TGF-β and with WT and Mut oligonucleotide) are those that bind non-specifically to the beads. Proteins present only after purification with TGF-β-induced extract and WT oligonucleotide are specific components of the "target complex".

components. They are also analysed by SDS-PAGE with silver staining of resolved proteins and mass spectrometry to identify the factors present in stained bands.

4.5 Development of a HaCaT cell line stably expressing HA-Smad3

An integral component of the purification system described above is a cell line stably expressing HA-Smad3. This is necessary to allow purification on a large scale and allows elution by peptide competition. HaCaT cells were selected as the parental line for generation of the necessary clones since this is the cell type in which the target complex has been best characterised (see Chapter 3 above). HaCaT cells are an immortalised but nontransformed human keratinocyte cell line of male origin. They are diploid, produce large amounts of extracellular matrix, have an epithelial morphology and grow as a monolayer in culture.

HaCaT cells were transfected with a plasmid carrying *HA-Smad3* cDNA inserted in-frame downstream of a promoter derived from the human *EF-1α* gene. The pSuperRetro vector, which carries a puromycin resistance gene driven by the human PGK promoter, was co-transfected at nine-fold lower levels. Puromycin is a fungally-derived antibiotic that inhibits the peptidyl transferase activity of both prokaryotic and eukaryotic ribosomes and the resistance gene carried by pSuperRetro encodes a puromycin acetyltransferase that inactivates this compound. Following transfection, cells were selected in puromycin and clones resistant to the drug were isolated. Since the HA-Smad3 vector was transfected at much higher levels, it was expected that many of these clones would express HA-Smad3. However, it was necessary to choose clones in which the exogenous Smad is functional and in which it is expressed at levels similar to those of the endogenous protein. Appropriate expression levels are important since highly overexpressed Smad3 could undergo non-physiological interations that might make it difficult to identify genuine partners by purification.

HA-Smad3 expression in a range of clones was determined by immunoflouresence and immunoblotting with antibodies specific to the HA tag (data not shown). Expression appeared to be generally low – probably reflecting a toxic effect of higher

levels of Smad3, a tumour suppressor that can contribute to cell cycle arrest and apoptosis.

I selected one clone for more extensive analysis of HA-Smad3 function. Immunoblot analysis of this clone with an antibody specific to Smad3 revealed that, when electrophoresis gels are allowed to run for a long time, HA-Smad3 can be resolved from endogenous Smad3 and is expressed at levels approximately equal to the endogenous (data not shown). Anti-HA immunoblotting was also repeated to confirm that HA-Smad3 expression in this line is stable (Fig. 4.3A) and anti-HA immunoprecipitation followed by immunoblotting with an antibody specific to phosphorylated Smad3 was employed to confirm that the stably transfected protein is phosphorylated in response to TGF-β (Fig. 4.3B). Similarly, immunostaining with an antibody specific to the HA tag confirmed stable expression and showed that – like endogenous Smads – HA-Smad3 undergoes nuclear accumulation in response to TGF-β (Fig. 4.3C).

These results showed that HA-Smad3 is generally functional but it was essential to show that it is also incorporated into the TGF-β-dependent complex that forms on the *c-Jun* promoter. This was achieved by EMSA with this element as probe and supershifts with antibodies specific to Smad3, Smad4 and the HA tag. Anti-HA supershifts revealed that HA-Smad3 is incorporated into these complexes and suggest that, under certain conditions at least, it may actually be incorporated preferentially to endogenous Smad3 (Fig. 4.4). Taken together with the findings described above, these data showed that a HaCaT cell line stably expressing fully functional HA-Smad3 had been generated.

4.6 Optimisation of two-step affinity purification

With the necessary cell line in hand, I proceeded to develop the purification system itself. The second step – DNAP – was optimised first since this involves fewer components than the IP step and since the physical conditions conducive to complex

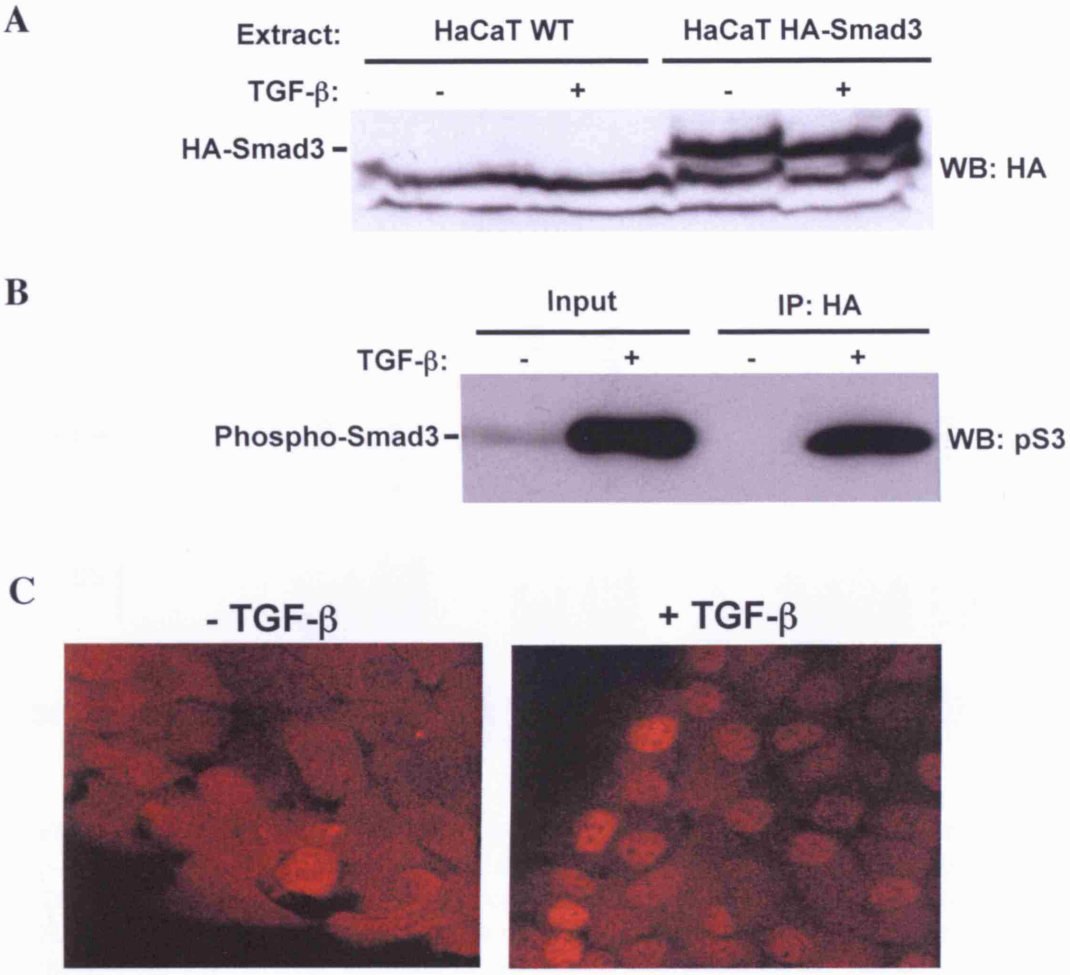


Fig. 4.3 A HaCaT cell line stably expressing haemaglutinin (HA)-tagged Smad3. *A) Western blot of wild-type (WT) HaCaT cells and HaCaTs stably transfected with HA-Smad3.* Whole-cell extracts were western blotted (WB) with an antibody specific to the HA tag. *B) Stably transfected HA-Smad3 is phosphorylated in response to TGF-β.* Nuclear extracts from TGF-β-induced and uninduced HaCaT cells stably expressing HA-Smad3 were immunoprecipitated with an antibody specific to the HA tag. Immunoprecipitates and crude extracts were then western blotted with an antibody specific to Smad3 that has been phosphorylated on its C-terminal SSVS motif. Input lanes were loaded with 50% total protein compared to immunoprecipitation inputs. *C) HA-Smad3 undergoes nuclear accumulation in response to TGF-β.* HaCaT cells stably expressing HA-Smad3 were immunostained with an antibody specific to the HA tag and visualised by laser-scanning confocal light microscopy.

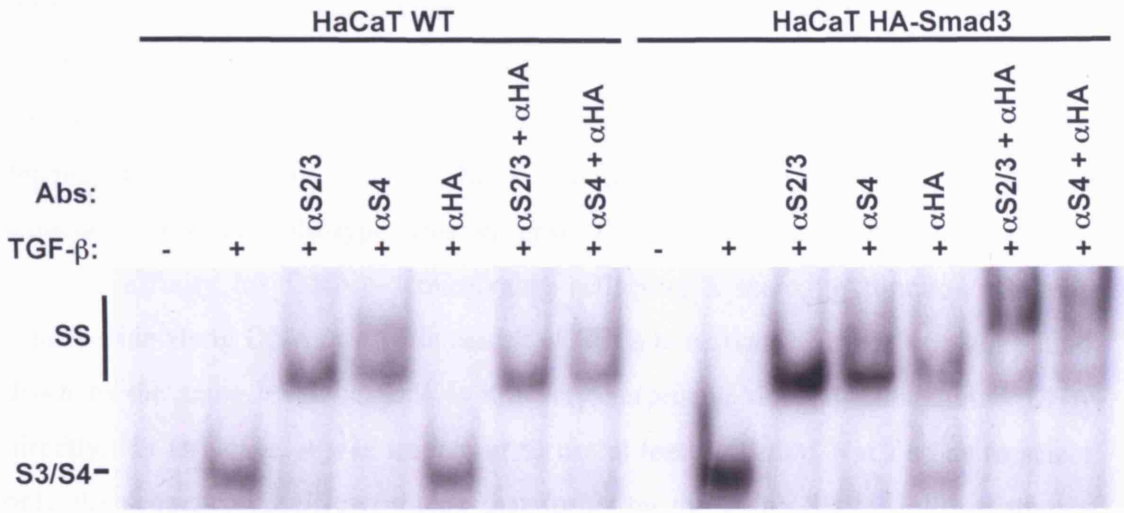


Fig. 4.4 Stably expressed HA-Smad3 is incorporated into Smad3-Smad4 complexes on the *c-Jun* SBR. Electrophoretic mobility shift assay was carried out with nuclear extract from TGF-β-induced and uninduced HaCaT cells. Wild-type (WT) cells and cells stably expressing HA-Smad3 were used. Antibodies (Abs) specific to Smad2/Smad3 (αS2/3), Smad4 (αS4) and the HA tag (αHA) were included as indicated. The positions the TGF-β-induced Smad3-Smad4 complex (S3/S4) and of antibody-induced supershifts of this complex (SS) are indicated.

formation in DNAP should also suit complex formation on the oligonucleotide included in the IP.

The DNAP procedure had already been employed for analytical purposes (see chapter 3 above) and, in this role, had already been optimised for NaCl concentration. DNAP was carried out with wild-type and mutant *c-Jun* SBR as bait. NaCl levels of 100, 120 and 140 mM were used and eluates were analysed by immunoblotting with antibodies specific to Smad2/Smad3 and Smad4. It was found that higher NaCl concentrations had little effect on the wild-type probe but that, on the mutant probe, low NaCl concentrations allowed Smad3 and Smad4 to be pulled-down in a TGF-β-dependent fashion (Fig. 4.5A). Since the mutant probe is highly divergent is sequence form the wild-type, this suggests that active Smad complexes have an intrinsic affinity for DNA – conceivably reflecting a scanning function whereby Smads slide along DNA *in vivo* in search of SBEs to activate. Smad2 was not pulled down to the same level but this is not very surprising since it cannot bind DNA directly. In any case, it was important to use at least 140 mM NaCl so as to select only the sequence-specific complex that forms on the *c-Jun* SBR. Further studies were carried out with higher NaCl concentrations and revealed that, beyond 140 mM NaCl, yields of Smad3 and Smad4 decline, dropping to zero above 200 mM (data not shown). Since a slight increase in the NaCl concentration above 140 mM would have little effect on background, I decided to adopt 140 mM as the purification condition.

Further optimisation was required to convert the DNAP from an analytical to a preparative technique. Specifically, due to cost constraints, it was necessary to separately titrate down all of the main reaction components to determine the minimum amounts compatible with maximal yield of Smad3 and Smad4. This was achieved by using different amounts of each component in DNAP and analysing the eluates by immunoblotting for Smad2/Smad3 and Smad4. The quantities of each component that were selected for an optimal DNAP with 200 μg of nuclear extract are as follows: beads (25 μg), biotinlyated oligonucleotide (1 μg) and mutant competitor oligonucleotide (5 μg) (Fig. 4.5, panels B-D).

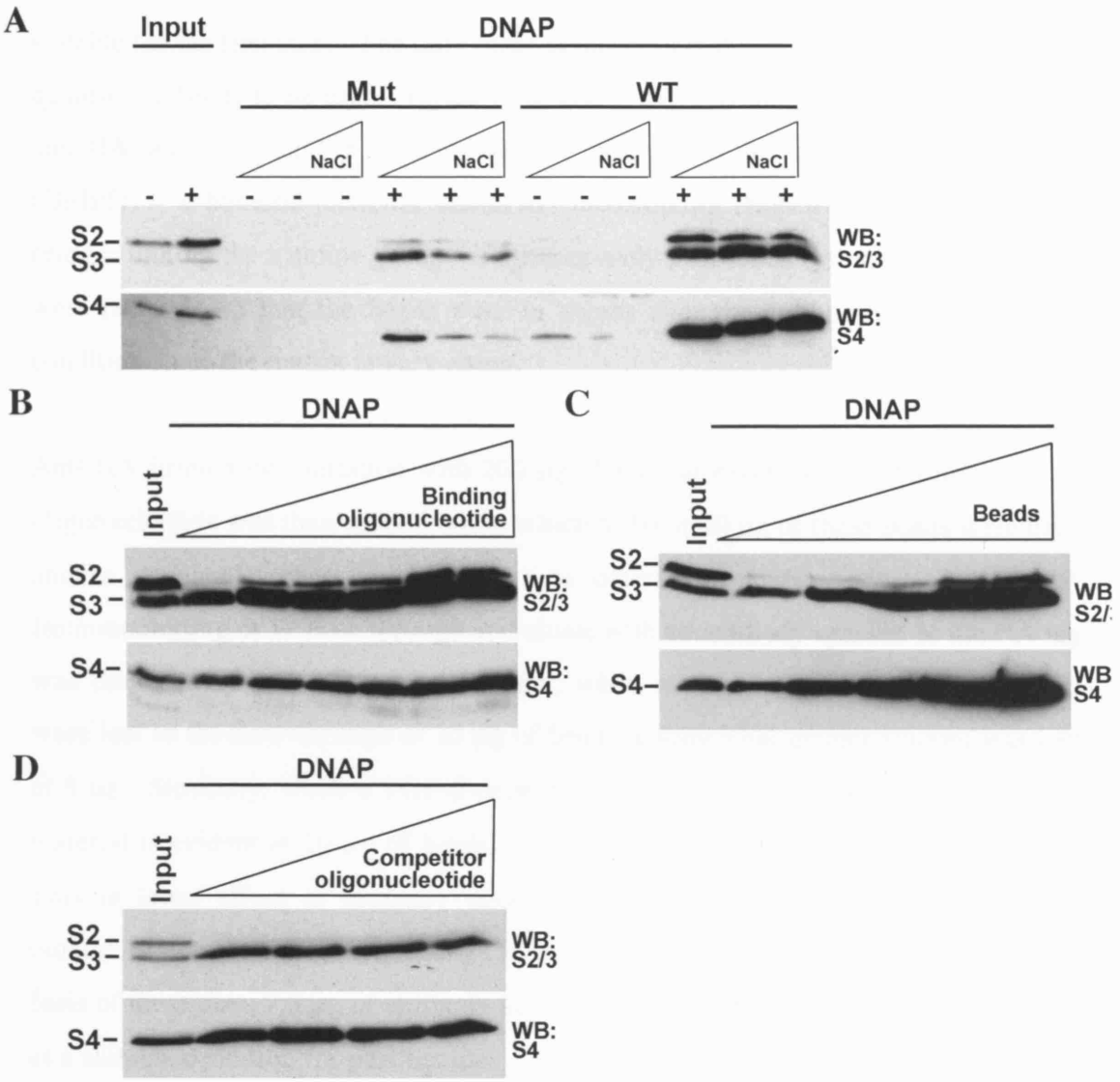


Fig. 4.5 Optimisation of DNA pull-down (DNAP). Nuclear extracts from TGF-β-induced wild-type HaCaT cells (panel A) and HaCaT cells stably expressing HA-Smad3 (panels B-D) were subject to DNAP as described in Fig. 3.5A. DNAPs were carried out with wild-type oligonucleotide bait only, washed and boiled in SDS sample buffer. Input lanes were loaded with 10% total protein compared to DNAP inputs, which were 200 μg in all cases. *A) Titration of NaCl.* NaCl concentrations used were 100, 120 and 140 mM. Binding oligonucleotide, competitor oligonucleotide (mutant in the SBE and factor Y binding site and not biotinylated) and neutravidin beads were all present in large excess. *B) Titration of binding oligonucleotide.* Quantities of 0.5, 1, 2, 4, 8 and 16 μg were used. Other components were in large excess. NaCl was at 140 mM. *C) Titration of beads.* Quantities of 6, 12, 25, 50 and 100 μg were used. Other components were in large excess. NaCl was at 140 mM. *D) Titration of competitor oligonucleotide.* Quantities of 5, 10, 15 and 20 μg were used. Other components were in large excess. NaCl was at 140 mM.

The conditions of NaCl, biotinylated oligonucleotide, and competitor oligonucleotide determined as optimal for the second purification step were assumed to also be suitable for the first step. The only other component that had to be optimised was the quantity of beads to be used. Beads were generated to which the mouse monoclonal anti-HA antibody 12CA5 was covalently conjugated by dimethyl pimelimidate (DMPD) – a homobifunctional imidoester crosslinking reagent that forms covalent bridges linking the ε amino groups of appropriately positioned lysine residues. These were prepared so that the beads were in excess over the antibody – an acceptable condition since the matrix is very cheap.

Anti-HA Immunoprecipitation with 200 μg of nuclear extract and in the presence of oligonucleotide was then carried out in which 5, 10 or 20 μg of these beads were used and in which all other variables were as determined to be optimal for DNAP. Immunoblotting of IP flow through and eluate with an antibody specific to the HA tag was then carried out. This revealed that, while negligible amounts of HA-Smad3 were lost to the flow-through at 10 μg of beads, a somewhat greater amount was lost at 5 μg. Similarly, while a TGF-β-dependent increase in the amount of precipitated material is evident at 10 μg of beads, no such increase is evident at 5 μg (Fig. 4.6). This is latter effect is probably because the quantity of HA-Smad3 present in uninduced nuclear extract is sufficient to saturate 5 μg of beads but not 10 μg. On the basis of these data, 7.5 μg of antibody beads per 200 μg of nuclear extract was chosen as a suitable condition for purification.

Further variables that required optimisation were the duration of each incubation, wash conditions and elution conditions. For both steps, it was found that incubations of anywhere between 4 and 12 hours gave equivalent performance (data not shown) – suggesting that, once complexes are formed, they are stable and no process of degradation or destabilisation is operating. Similarly, washing of both steps in column format with 50 bed-volumes of 140 mM NaCl buffer did not to lead to serious loss of complex components (was done for the experiment shown in Fig. 4.8). In terms of elution, five sequential elutions with 300 μg ml^{-1} HA peptide were found to fully elute the first step (data not shown) and incubation for one hour with buffer

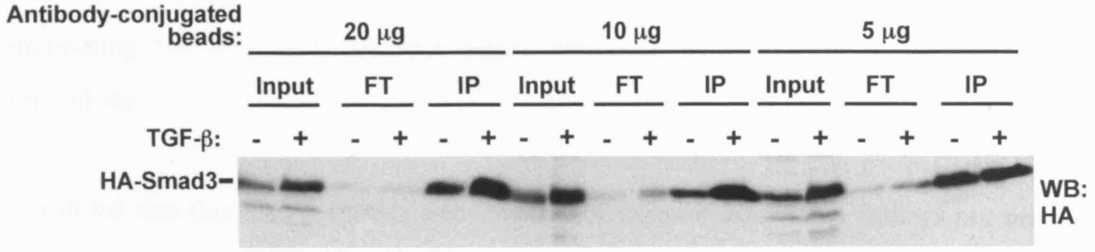


Fig. 4.6 Optimisation of step 1: Titration of antibody-conjugated beads. Beads were conjugated to anti-HA antibody (data not shown) and the indicated quantities were used in IP. IP was carried out as in Fig. 3.1A with 200 μg of nuclear extract from TGF-β-induced HaCaT cells stably expressing HA-Smad3. All other variables were taken as having the same optimum as in DNAP. After incubation, supernatant (flow through, FT) was recovered and beads were washed and boiled in SDS sample buffer. FT was concentrated by ultrafiltration and both FT and IP eluates were analysed by western blotting (WB) with an antibody specific to the HA tag. Inputs were 10% total protein compared to IP input for a single immunoblot lane.

containing 600 mM NaCl had already been shown to efficiently elute the second step (Fig 4.1A, Fig. 4.8).

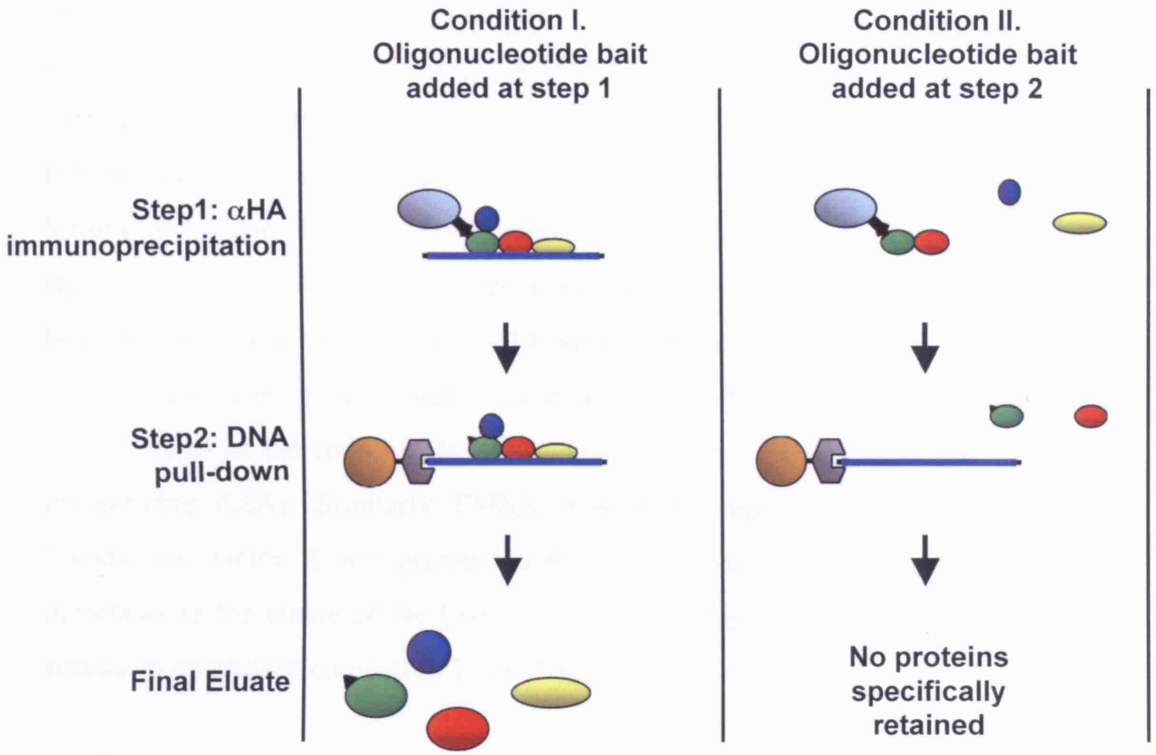
4.7 Implementation of the purification and its use to confirm the presence of additional factors

Having optimised all the core variables of the purification, it was necessary to implement the entire process and evaluate its performance. This was carried out by incubating the first step for four hours, followed by overnight incubation of the second step.

I realised that this also provides a means to determine if additional factors are present in the complex. If factors X and Y, as well as enhancing the interaction of Smad3 and Smad4 with DNA, only associate with the Smads on DNA then retention of the target complex in the purification should be completely dependent on the inclusion of oligonucleotide in the first step.

In order to test the integrated purification and confirm that there are non-Smad factors present in the complex that are only incorporated on DNA and are also essential for DNA binding, I designed an experiment in which the two purification steps were combined as described above. This was done under two different conditions. In condition I, biotinlyated oligonucleotide was present throughout the purification. In condition II, it was omitted in the first step and added only at the beginning of the second step (Fig. 4.7).

The predicted result was that, in condition I, EMSA and immunoblotting would show that Smad3, Smad4 and all other factors essential for complex formation are present in the final eluate. In condition II, however, it was predicted that Smad3-Smad4 heterodimers would be precipitated in the first step but that, due to the absence of oligonucleotide, factors X and Y would flow through. This would prevent HA-Smad3 and Smad4 from binding DNA in the second step, with the result that they



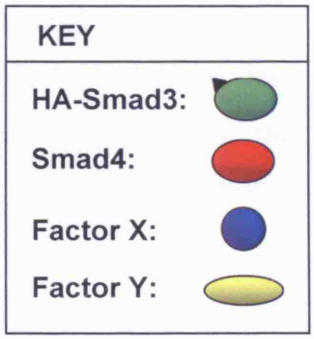


Fig. 4.7 Predicted DNA-dependence of two-step affinity purification. If two-step purification is carried as described in Fig. 4.2, all essential components of the complex are expected to be present in the final eluate (condition I above). However, if oligonucleotide bait is omitted from the first step (condition II above), it is expected that a Smad3-Smad4 dimer will be present in the eluate of the first step, but will be unable to bind the oligonucleotide when it is provided at the beginning of the second step. This is because factors X and Y are thought to associate with the Smad dimer in a DNA-dependent fashion. It is also predicted that, in the eluate of the first step, less Smad4 will be present in condition II than in condition I. This is because of the potentiation of the Smad3-Smad4 interaction that is observed on the *c-Jun* SBR.

would flow through (Fig. 4.7). EMSA and immunoblotting would thus show an absence of HA-Smad3, Smad4 or complex formation in the final eluate.

When the experiment is actually carried out with 3 mg of nuclear extract from TGF-β-induced HaCaT cells stably expressing HA-Smad3, the results that are obtained correspond closely to what was expected. Immunoblotting shows that, in condition I, HA-Smad3 is pulled down very efficiently in the first step. Substantial amounts of Smad4 are co-precipitated, but a significant amount also flows through, which represents Smad4 molecules that are associated with partners other than HA-Smad3. In the second step, the amount of HA-Smad3 that was in excess relative to Smad4 after the first step flows through. No Smad4 is detectable in the flow through of the second step. In the final eluate, substantial amounts of HA-Smad3 and Smad4 are present (Fig. 4.8A). Similarly, EMSA shows that complexes containing HA-Smad3, Smad4 and factor X are present in the final eluate. These complexes are not detectable in the eluate of the first step since the oligonucleotide bait is present and acts as an unlabelled competitor in the EMSA (Fig. 4.8B).

In condition II, immunoblotting shows that, as in condition I, HA-Smad3 is efficiently precipitated in the first step. However, due to the absence of oligonucleotide, co-precipitation of Smad4 is much less efficient than in condition I. In the second step, all HA-Smad3 and Smad4 present after the first step flow through (Fig 4.8A). In the EMSA, no TGF-β-dependent complex formation is detected on a probe corresponding to the *c-Jun* SBR (Fig. 4.8B).

These results can be interpreted in two ways. One is that, because the lack of wild-type oligonucleotide in the first step in condition II, Smad4 is co-precipitated less efficiently than in condition I. The result is that fewer Smad3-Smad4 complexes are present to bind DNA in the second step and no complex components are retained to detectable levels in the final eluate. This, however, cannot be the case. In condition II, no unlabelled DNA is present after the first step, HA-Smad3 is abundant and Smad4 is present to levels at least equal to those found in the final eluate of condition I. Yet no specific complex is detected by EMSA.

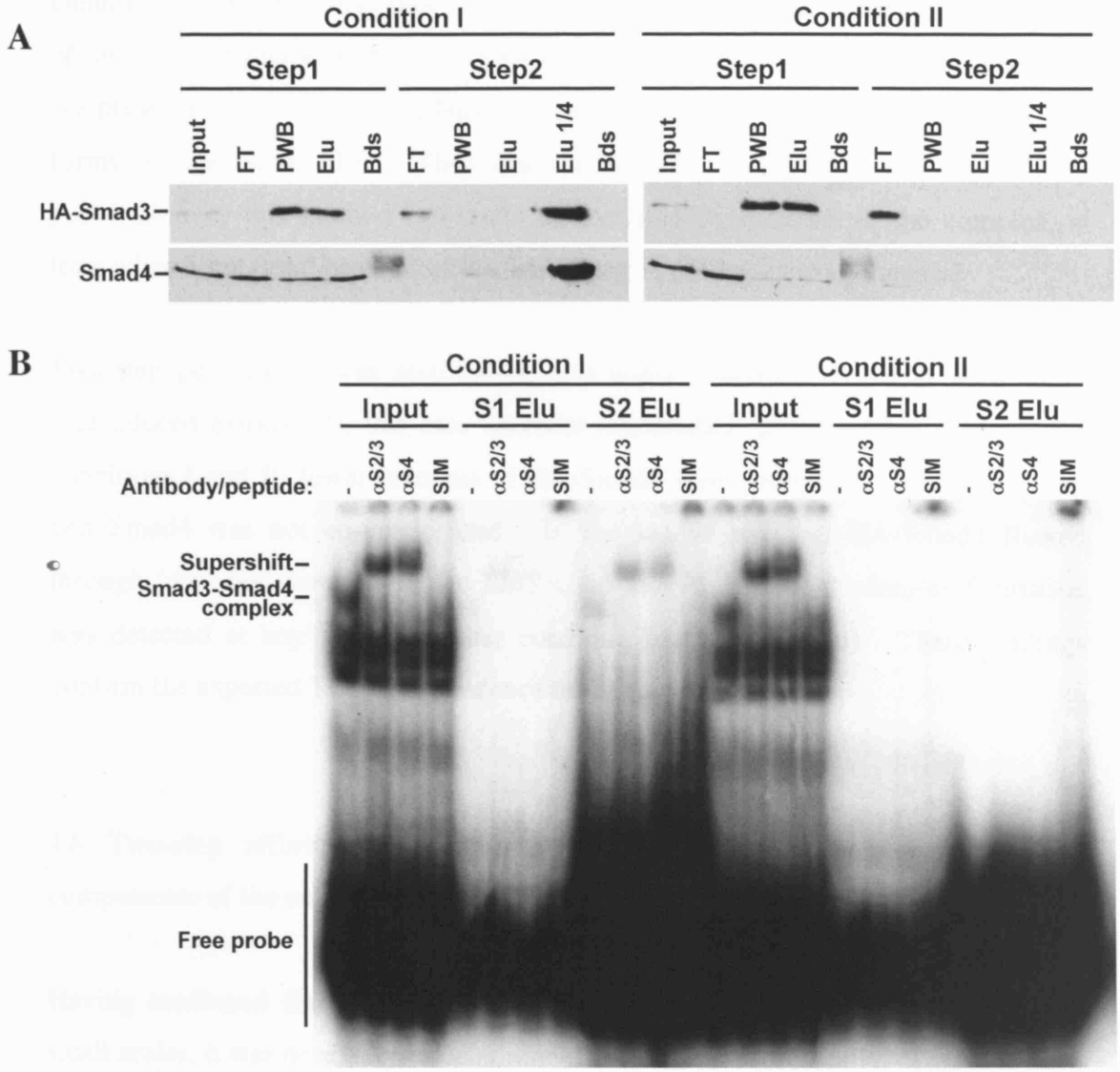


Fig. 4.8 DNA-dependence of two-step affinity purification confirms the presence of other factors in addition to Smad3 and Smad4. Conditions I and II are as described in Fig 4.6. All experiments were done with nuclear extracts from TGF-β-induced HaCaT cells stably expressing HA-Smad3. *A) Samples from two-step purification analysed by western blot.* 0.5 % samples were taken of the input for step 1 and of the flow-through (FT) for both steps. 1 % samples were taken of pre-wash beads (PWB), eluate (Elu) and final beads (Bds) for both steps. 25% samples were taken of the eluate for the second step (Elu 1/4). All of these samples were western blotted (WB) with antibodies specific to Smad4 (S4) and the HA tag (HA), as indicated. *B) Samples from two-step purification analysed by EMSA.* Samples were taken at the indicated stages and combined with radiolabelled probe corresponding to the *c-Jun* SBR. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and peptide corresponding to the Mixer SIM were added, as indicated. The positions of free probe, of the TGF-β-induced Smad3-Smad4 complex and of antibody-induced supershifts of this complex are indicated. Note that complex is not detected in the eluate of step 1 in condition I because the oligonucleotide bait is present and competes with the labelled probe.

The more likely explanation, is that the complex contains at least one non-Smad factor that is only incorporated on DNA and which is essential for DNA binding. In condition II these factor(s) flow through the first step, preventing further association of Smad3 or Smad4 with the *c-Jun* SBR in EMSA or DNAP. These data thus confirm the presence of at least one additional protein in the TGF-β-dependent complex that forms on the *c-Jun* SBR. They also demonstrate clearly that two-step affinity purification by this method can retain all essential components of the complex, at least when 3 mg (total protein) of nuclear extract is used as starting material.

Two-step purification was also carried out under conditions I and II with TGF-β uninduced extract. In this case identical immunoblot patterns were obtained for conditions I and II: lower amounts of HA-Smad3 were precipitated in the first step and Smad4 was not co-precipitated. In the second step, all HA-Smad3 flowed through (data not shown). In the EMSA, no TGF-β-dependent complex formation was detected at any stage in either condition (data not shown). These findings confirm the expected TGF-β-dependence of the purification.

4.8 Two-step affinity purification on a 30 mg scale retains all essential components of the complex but does not permit subunit identification

Having confirmed that the purification system described performs as predicted on small scales, it was necessary to determine if it scales up successfully. This is often a problem in preparative procedures. 30 mg (total protein) of nuclear extract from TGF-β-induced HaCaT cells stably expressing HA-Smad3 were subjected to two-step purification as described above. Uninduced extract was not used since the TGF-β-dependence of two-step purification had already been established (see section 4.7 above). Samples taken at various stages were analysed by immunoblotting with antibodies specific to Smad4 and the HA tag and by EMSA with a radiolabelled probe corresponding to the *c-Jun* SBR.

Immunoblotting showed that significant amounts of HA-Smad3 and Smad4 are present in the final eluate. HA-Smad3 appears to be retained strongly throughout the

purification. Anomalously, the levels of Smad4 present in the final eluate appear to be higher than in the eluate of the first step (Fig. 4.9A). However, since the levels being measured are close to the detection threshold this may simply be due to assay variation.

Similarly, the EMSA showed that additional components of the complex are also present in the final eluate. However, these appear to be present at very low levels since complex formation is detected only when stabilised with antibodies specific to Smad2/Smad3 or Smad4. A further anomaly is that complex formation is also detected in the eluate of the first step (Fig. 4.9B).

Despite technical problems, these results demonstrate that it is possible to retain HA-Smad3, Smad4 and other components of the complex through two affinity steps on a 30 mg scale. However, analysis by SDS-PAGE with silver staining of resolved proteins revealed no TGF-β-dependent bands (data not shown). Further optimisation will be needed to increase the yield to the point where complex components are detectable by silver staining and identifiable by mass spectrometry. This is likely to involve further scale-up of the purification, perhaps to a scale of several hundred milligrams.

4.9 Conclusions

This chapter provides clear evidence that there is at least one essential non-Smad component in the TGF-β-dependent complex that forms on the *c-Jun* SBR. It also describes the development of a purification system that can retain substantial amounts of HA-Smad3, Smad4 and all other essential components of the complex through two affinity steps. This provides a system by which factor X, factor Y and any other components present can be isolated. This, in turn, will allow physical and functional characterisation of their role in TGF-β signalling and should generate significant new insights into Smad-transcription factor interaction. However, before these goals are realised, further work will be required to scale-up the purification system and

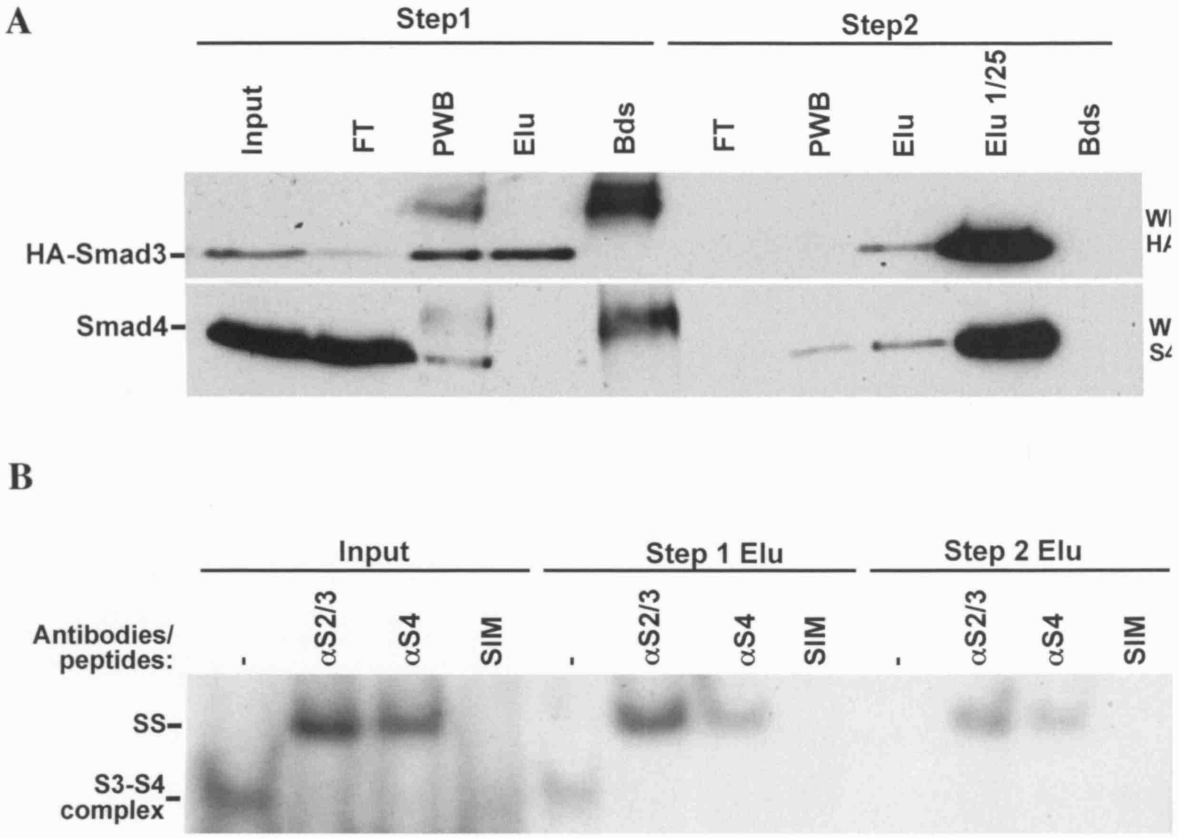


Fig. 4.9 Two-step purification on a 30 mg scale retains Smad3, Smad4 and all other factors required for TGF-β-dependent complex formation on the *c-Jun* SBR. Purification was carried out with wild-type and mutant oligonucleotide bait and TGF-β-induced nuclear extract from HaCaT cells stably expressing HA-Smad3. *A) Samples from two-step purification analysed by western blot.* 0.1% samples were taken at the indicated stages: input, flow-through (FT), pre-wash beads (PWB), eluate (Elu) and final beads (Bds). For step 2 eluates, 4% samples were also taken (Elu 1/25). Samples were then western blotted (WB) with antibodies specific to Smad4 (S4) and the HA tag (HA). *B) Samples from two-step purification analysed by EMSA.* Samples were taken at the indicated stages and combined with radiolabelled probe corresponding to the *c-Jun* SBR. Antibodies specific to Smad2/Smad3 (αS2/3) and Smad4 (αS4) and peptide corresponding to the Mixer SIM were added, as indicated. The positions of the TGF-β-induced Smad3-Smad4 (S3-S4) complex and of antibody-induced supershifts of this complex are indicated. The complex present in step 2 eluate is very weak and is only detected when stabilised by antibody supershifts. Anomalously, complex is detected in the eluate of step 1. This was not expected, since the oligonucleotide bait acts as a competitor and may reflect poor retention of this bait.

optimise it on a larger scale. Only then will it be possible to isolate sufficient quantities of each complex component to allow identification by mass spectrometry.

CHAPTER 5

DISCUSSION

5.1 Summary of results

Previous studies identified a TGF-β-dependent complex that forms on the *c-Jun* SBR and which contains Smad3, Smad4 and at least two unidentified proteins, temporarily designated factors X and Y (Fig. 1.9B) (Inman and Hill, 2002; Wong et al., 1999). The work described in this thesis confirms the presence of non-Smad components in this complex and provides a purification system that – with some further optimisation – should allow identification of these proteins. The detection of factors X and Y created an opportunity to explore Smad-transcription factor interactions in fully differentiated cells of the mammalian adult and, by making it possible to identify these proteins, my work provides the means to exploit this opportunity. A second result of my work is the finding that Smad4 and SnoN associate with the *c-Jun* SBR *in vitro* in the absence of TGF-β. It is possible that SnoN functions to repress *c-Jun* in this context but the function of Smad4 on the *c-Jun* SBR in the absence of TGF-β is not clear.

5.2 Smad4 is essential for TGF-β-dependent complex formation on the *c-Jun* SBR

The work of Inman and Hill had shown that the TGF-β-dependent complex that forms on the *c-Jun* SBR contains one molecule of Smad3 and one molecule of Smad4 (Inman and Hill, 2002). These are likely to be the core components of the complex and are responsible for its TGF-β dependence.

It is an established principle that the assembly of functional transcription factor complexes on promoters and enhancers is often highly dependent on the availability of multiple distinct components of the complex – if any component is not present, the complex does not form and the gene is not induced (Fig 1.1) (Courey, 2001). This

seems to be true also of the TGF-β-dependent complex that forms on the *c-Jun* SBR. Removal of factor X or factor Y has a destabilising effect (see section 5.3 below) and – since they are the core components of the complex – it seemed likely that loss of Smads would have an even greater effect. I tested this hypothesis in the case of Smad4 using a cell line in which the expression of this protein can be knocked down in a tetracycline-inducible fashion. Treatment of these cells with tetracycline showed that complex formation is indeed greatly diminished when Smad4 is depleted, supporting the model that Smad4 is required for complex formation (Fig. 3.2). The same experiment was not conducted for Smad3 but it seems likely that a similar conclusion would be reached. It is, however, possible that the effect of Smad3 knock-down would show a different dose-dependency – reflecting the fact that Smad3 might be expressed to different levels or incorporated with different efficiency.

Despite the necessity of Smad4 for complex formation *in* vitro, it has been shown that, when Smad4 is knocked-down, the induction of *c-Jun* by TGF-β is unaffected (Levy and Hill, 2005). This could be because only low levels of complex formation on the SBR are required for *c-Jun* induction or because TGF-β can induce *c-Jun* by two redundant pathways: a Smad4-dependent pathway acting through the SBR and another pathway acting through a separate element. Since Smad4 knock-down was very efficient, the second possibility seems much more likely. It has also been shown that, in luciferase assays, an intact SBR is required for TGF-β-dependent activation of the *c-Jun* promoter (Wong et al., 1999). However, it is possible that a redundant Smad4-independent pathway can activate *c-Jun* through an element located outside the region analysed in these assays.

As well as showing that Smad4 is required for complex formation, I also showed that the interaction between Smad3 and Smad4 can be potentiated on the *c-Jun* SBR. This was achieved by immunoprecipitating Smad3 and HA-Smad3 in the presence of wild-type or mutant oligonucleotide corresponding to the *c-Jun* SBR. Under appropriate conditions, Smad4 is co-precipitated with Smad3 or HA-Smad3 only weakly but this interaction is strongly potentiated by inclusion of the wild-type oligonucleotide, indicating that Smad complexes are stabilised on DNA (Fig. 3.1).

5.3 Additional factors are present in the TGF-β-dependent complex that forms on the *c-Jun* SBR

Inman and Hill also showed that there are probably two non-Smad factors present in the TGF-β-dependent complex that forms on the *c-Jun* SBR. Factor X was identified by competition studies with SIM peptide and factor Y by truncation of the DNA binding site (Inman and Hill, 2002). I replicated these findings and thus confirmed the presence of factors X and Y in the complex (Fig, 3.3). A number of other experiments also confirmed the presence of additional factors.

One way in which this was achieved was two-step purification carried out either with wild-type *c-Jun* SBR oligonucleotide present throughout or present only in the second step and omitted in the first step (Fig. 4.7). In this experiment, Smad3, Smad4 and all other essential components of the complex are retained in the final eluate when wild-type oligonucleotide is included in both steps but not when it is omitted in the first step (Fig. 4.8). This indicates that there must be at least one additional factor present in the complex, that it only associates with the Smads on DNA and that it is required for their association with the *c-Jun* SBR. When wild-type oligonucleotide bait is present throughout the purification, the additional factor(s) can associate with the Smads on the oligonucleotide in the first step and remain associated throughout the purification. However, when oligonucleotide is omitted in the first step, these factor(s) cannot associate with the Smads and flow through this step. As a consequence, only Smad3-Smad4 dimers are pulled down in the first step. When oligonucleotide bait is then provided in the second step, the Smads cannot bind to it and instead flow through the purification.

Likewise, I confirmed the existence of factor Y by mutational analysis of its binding site. A series of radiolabelled probes were made in which successive 2-basepair units of the GCCA triplet found at this position were changed in sequence and used in EMSA analysis (Fig. 3.10). This revealed that the second and third repeats of the GCCA triplet are required for optimal complex formation. The first repeat makes only a slight contribution and sequences downstream of the GCCA triplet are dispensable (Figs. 3.10 and 3.11). This is an important finding, since in the original

studies of Inman and Hill, all three repeats were mutated together, without changing them individually (Inman and Hill, 2002). Their work left open the possibility that only the first repeat – which is closest to the SBE – is important for Smad3/Smad4 binding and that it forms part of the site bound directly by the Smads. This would be consistent with the known affinity of Smads for GC rich regions (see chapter 1 above). However, the finding that the first repeat is not important, together with the mapping of a detailed binding site indicates that this region is not bound by a Smad and thus supports the existence of factor Y.

Prior to actual identification of all its components, there are a number of other studies that could be conducted to characterise the complex in greater detail. To further confirm the existence of factor X, epitope-tagged mutants of Smad3 with sequence changes in the SIM-binding pocket could be made, using previous mutagenesis of Smad2 as a guide (Randall et al., 2002). These could then be used in EMSA to confirm that this region of Smad3 is important for complex formation. Such experiments would complement the SIM competition studies already carried out and it would be expected that, in further studies with SIM peptide, these mutants would be insensitive to competition. The DNA site bound by factor Y could also be characterised in more detail (see section 5.4 below).

5.4 The minimal *c-Jun* SBR is a 32-basepair region that contains a Smad-binding element plus a site bound by factor Y

As well as demonstrating the existence of additional factors in the TGF-β-dependent complex that forms on the *c-Jun* SBR, I also mapped in detail the DNA sequences that are required for its formation. As already described (see section 5.2 above) this involved mutagenesis to show that the factor Y binding site is a GCCA double repeat downstream of the SBE. Similarly, truncation studies showed that the minimal *c-Jun* SBR is an element of around 32-basepairs. From 5' to 3' this consists of 5 appparently non-specific basepairs, the AGAC triplet bound by the Smads, a spacer of 4 basepairs and the GCCA triplet. Complex formation on this element is approximately 6-fold lower than on the original 55-basepair probe (Fig 3.9B). On

probes where one or the other end of the 55-basepair element (but not both) is truncated to a greater extent than in the 32-basepair probe, complex formation is reduced to very low levels (Fig. 3.7B and Fig. 3.8B). This indicates that the 32-basepair element corresponds closely to the minimal *c-Jun* SBR.

The DNA binding site could be further characterised by DNaseI footprinting. In this method, double-stranded probes corresponding to the binding site of interest are synthesised and one strand is end-labelled with a radioactive tracer. These are then incubated with extract and degraded with DNaseI under conditions where, on average, one double-stranded cut is made per probe duplex. The reaction mixture is then denatured in urea to remove bound protein and separate the DNA strands, which are analysed by polyacrylamide gel electrophoresis and autoradiography. This reveals the regions of the probe protected from degradation by bound protein. If DNaseI footprinting were carried out with the *c-Jun* SBR as radiolabelled probe and extract from TGF-β-induced and uninduced HaCaT cells, the expected result from the TGF-β-uninduced condition would be a ladder of bands corresponding to cleavage at every position in the probe. However, in the TGF-β-induced condition, a space would be expected in this ladder corresponding to the region protected by the Smads and factor Y. This should confirm the extent of the factor Y binding site as determined by mutagenesis and would thus provide further confirmation of the presence of factor Y.

Similarly, successful *in vitro* site selection would reveal the physically optimal binding site for factor Y and could lead to identification of this factor on the basis of known DNA sites for transcription factor binding. As described in section 3.10, this was unsuccessful because of non-specific interactions between abundant DNA-binding proteins and the beads. This problem might be solved in future by carrying out the first round under more stringent conditions that restrict the formation of background complexes. Measures to improve stringency could include inclusion of a larger excess of unrelated double-stranded DNA to titrate DNA-binding proteins responsible for the background (polydeoxyinosinate-deoxycytidine was used in these selections). Similarly, extracts could be pre-cleared using beads on which oligonucleotides of this kind are immobilised. Other measures could include raising

the NaCl concentration of the immunoprecipitations or including additional detergents (1% triton X-100 was used in these studies).

5.5 Two-step affinity purification retains all factors required for TGF-β-dependent complex formation on the *c-Jun* SBR

Although mutagenesis of the SIM-binding pocket of Smad3 and DNaseI footprinting of the DNA site would give some further insight into TGF-β-dependent complex formation on the *c-Jun* SBR, little further progress can be made until factors X and Y are identified and characterised. In this thesis I describe a two-step affinity purification system designed to isolate all components of the complex (Fig. 4.2).

Starting material for purification was derived from a HaCaT cell line expressing HA-Smad3 at levels approximately equal to endogenous Smad3. HA-tagged Smad3 was used since anti-HA antibodies are cheap and readily available and since this system allows competitive elution with HA peptide. HaCaTs were chosen to provide starting material for this system since the TGF-β pathway has been well characterised in this context and the target complex is known to form in extract from HaCaT cells. Near-endogenous expression levels were also important since higher levels of expression might lead to non-physiological interactions that would complicate interpretation of purification data.

The purification itself involved immunoprecipitation of the complex with an anti-HA antibody in the presence of a biotinylated oligonucleotide corresponding to the DNA binding site (Fig. 3.1A). The presence of this oligonucleotide stabilises the complex and allows eluate from the first step to be used directly in the second step, where the biotin moiety is captured on a neutravidin matrix (Fig. 3.5A). This system was adopted because single-step purification by DNA pull-down was found to bring down a number of background proteins too high to allow identification of specific complex components.

For a given quantity of extract, optimal concentrations of NaCl, binding oligonucleotide, mutant competitor oligonucleotide and beads were determined for each affinity step (Fig. 4.5). The incubation times and elution conditions of each step were also optimised (data not shown). Under these optimised conditions and with 3 milligrams total protein as starting material, the system was able to retain substantial amounts of Smad3 and Smad4, along with all other factors essential for complex formation on the *c-Jun* SBR (Fig. 4.8).

When scaled-up to 30 milligrams, the system also retained the target complex but was less efficient and had a number of anomalies in its performance (Fig. 4.9). In addition, when eluates were analysed by SDS-PAGE with silver staining of resolved proteins, no bands corresponding to specific complex components were detected (data not shown). This indicated the need for further optimisation of the system on this scale. Ways to achieve this might include moderate increases in the quantities of critical reaction components such as beads or binding oligo. The fact that complex formation is detectable in EMSA analysis of the first purification step (Fig. 4.9) suggests that levels of the oligonucleotide bait – which acts as an unlabelled competitor – are low and that increasing them somewhat might lead to an increase in performance. Other possibilities include slight adjustments of incubation times, NaCl concentrations, washing conditions or elution conditions. In any case, it is likely that several hundred milligrams of starting material would be required to allow identification of novel Smad partners.

If these measures are unsuccessful, a number of more radical adaptations of the system are possible. One option would be to attempt single-step purification based on the first step of the current system. This would retain all Smad3-interacting proteins but, if purification conditions with and without oligonucleotide bait were compared, it might be possible to identify DNA-dependent partners. Due to the need to employ low NaCl concentrations, this system would also suffer the same background problem as DNA pull-down (Fig. 4.1B). However, because the immunoprecipitation step is eluted by specific competition with HA peptide, the problem might be less severe than with DNA pull-down alone.

Another possibility would be to covalently crosslink complex components to DNA and/or to each other using formaldehyde or other means. This would enable washing at higher NaCl concentrations and might thus help alleviate the background problem – possibly allowing single-step purification. Such a system would probably involve crosslinking being carried out in a small volume in the absence of beads, followed by dilution and rapid capture of the complex on an anti-HA or neutravidin matrix. In this way, the problem of also crosslinking background proteins to the beads might be avoided. Together with washing at higher NaCl concentrations, this might be enough to overcome the background problem and transform the second step of the current system into a feasible single-step purification. If reversible crosslinking reagents such as dimethylsuberimidate (DMS) – whose effects can be reversed with dithiothreitol – were used, it might still be possible to measure formation of the target complex in the final eluate by EMSA. If formaldehyde were used, boiling in SDS sample buffer would suffice to reverse the crosslinks for analysis by SDS-PAGE with immunoblotting or with silver staining of resolved proteins but EMSA would be impossible.

Instead of direct capture on an affinity matrix, size exclusion chromatography could first be used to separate crosslinked protein-DNA complexes from other components of the extract. Free crosslinking reagent would also be removed and, while some complex might be lost through crosslinking to the matrix, enough should pass through for capture on an affinity matrix. In all of these alternative concepts, a DNA bait consisting of a concatomerised 32-basepair SBR might be more efficient than a single 55-basepair SBR. Although the shorter element is less efficient in EMSA (Fig 3.9), the presence of multiple elements in a concatomer should compensate – reflecting the fact that the flanking DNA that stabilises complex formation does not need to be of a specific sequence. And concatomerisation of the 32-basepair element, rather than the 55-basepair element, would allow more repeats to be included. This should increase the binding capacity of the matrix, allowing the use of fewer beads and thus reducing the number of binding sites for background proteins.

The two-step purification strategy I have developed should, with some further scale-up and optimisation, be sufficient to isolate specific complex components at levels

that are detectable by silver staining and identifiable by mass spectrometry. However, even if this did not prove to be the case, there are a number of derivations of the basic method that seem equally promising as tools to isolate all components of the complex. It is thus highly likely that the methods described in this thesis will enable identification of these proteins and allow new insights into Smad-transcription factor interactions.

5.6 Future directions in identifying and characterising proteins that reside in the TGF-β-dependent complex that forms on the *c-Jun* SBR

Upon identification of TGF-β-dependent components in the eluate of the two-step purification or any other system designed to isolate components of the target complex, it would be necessary to validate these as genuine Smad partners. This could be achieved by EMSA of the *c-Jun* SBR with supershifts of the identified factors, or by DNAP with immunoblotting for these proteins. If – as seems likely – antibodies against the endogenous proteins would need to be generated before this could be attempted, preliminary validation could be carried out with transfected epitope-tagged proteins. Immunoprecipitation in the absence of *c-Jun* SBR oligonucleotide could also be employed to determine if any of the identified factors interact with Smads in solution.

Once validated, the role of novel Smad partners in the TGF-β pathway could be investigated. This would begin with molecular analysis of their interaction with the Smads. For interactions that occur in solution, reconstitution with recombinant proteins could be used to show which interactions might be direct. Deletion mapping or overexpression of tagged fragments followed by immunoprecipitation could then be used to identify the interaction domains in both the Smads and their novel partners. Data obtained from such studies could be validated by peptide competition studies or by making specific point mutations. Similarly, SIM competition in EMSA could be used to identify which of the novel partners is factor X. Once factor X is identified, a key task would be to examine the region of this protein that interacts with the SIM-binding pocket of Smad3. It would be important to investigate the similarities and

differences between this interaction and that of Smad2 with the Mixer SIM – a goal that, as before, could be achieved by a combination of directed mutagenesis and molecular modelling (Randall et al., 2002).

As with factor X, it might be possible to identify which of the novel partners is factor Y by using probes mutated in the factor Y binding site. Deletion analysis and point mutations could then be used to identify the DNA binding domain of this protein. Even though the affinity is likely to be low (Inman and Hill, 2002) it could also be possible to analyse the interaction between factor Y and DNA by means of *in vitro* reconstitution with elevated concentrations of DNA and protein. This would be made easier if site selection studies identified an optimal binding site for factor Y.

Alongside the physical characterisation of novel partners, it would be possible to also analyse their biological function in the TGF-β pathway. Inducible shRNA knock-down with microarray analysis could be used to determine which TGF-β target genes require these factors for their induction and similar protein ablation studies could be used to determine which cellular responses to TGF-β are reliant on these factors. Processes to be investigated would include cell cycle arrest, apoptosis and EMT. A similar approach was used by Levy and Hill to investigate the role of Smad4 in TGF-β signalling (Levy and Hill, 2005). Expression of identified partners in primary cells and tumour samples could also be explored – probably by realtime RT-PCR – to obtain some indication of their function in tissue homeostasis and tumourigenesis.

It is also possible that newly identified Smad partners would work as a platform for integrating the TGF-β pathway with other signalling cascades. RNase protection, immunoblotting, mass spectrometry and immunocytochemistry could be used to determine if they undergo any change in expression/localisation or any covalent modification in response to extracellular signals. Similarly, DNAP and EMSA could be used to determine if their interaction with the Smads is regulated. The effects of any such cross-talk on the cellular functions of these partners could then be investigated. It could also be investigated whether TGF-β itself induces any change in the expression, localisation or activity of these proteins.

5.7 Smad4 and SnoN associate with the *c-Jun* SBR in the absence of TGF-β

A second set of findings described in this thesis are that Smad4 and SnoN associate with the *c-Jun* SBR in the absence of TGF-β. SnoN is unable to bind DNA directly but does not appear to be recruited via Smad4 either. Instead it continues to be recruited in cells where Smad4 has been depleted by inducible shRNA knock-down (Ross, SJ and Hill, CS, unpublished data). It thus seems likely that SnoN is recruited by another factor that binds the *c-Jun* SBR directly.

The role of Smad4 on the *c-Jun* SBR in the absence of TGF-β is difficult to predict. One possibility is that Smad4 functions as a "landing pad" to nucleate assembly of the activating complex. Other possibilities are that Smad4 functions to keep the promoter in a state that is competent for activation or that it functions to repress *c-Jun* in the absence of TGF-β. It has been shown by microarray that *c-Jun* induction by TGF-β is not affected by Smad4 knock-down and that Smad4 knock-down does not result in elevated *c-Jun* expression in the absence of TGF-β. This appears to rule out a role for Smad4 in repressing *c-Jun* in the absence of signal but, since there are thought to be two redundant mechanisms by which TGF-β activates *c-Jun*, both of the other possibilities remain open. However, with the data available, it is impossible to conclude which is the correct hypothesis.

The role of SnoN on the other hand is probably more straightforward, it is a corepressor that recruits the N-CoR–Sin3– HDAC complex (Stroschein et al., 1999) and it probably functions to repress *c-Jun*. This may be needed to prevent inadvertent activation of *c-Jun* by low levels of autocrine TGF-β signalling, a role that would be similar to the function of FKBP12 in preventing inadvertent receptor activation (Shi and Massague, 2003). This is consistent with the known role of SnoN as a general antagonist of Smad function (Luo, 2004). SnoN is rapidly degraded in response to exogenous TGF-β, an event that may be required for efficient induction of *c-Jun*. It is then re-synthesised after 2 hours of signalling (Bonni et al., 2001; Stroschein et al., 1999; Sun et al., 1999), and may associate again with the promoter to suppress transcription. This is consistent with the transient nature of *c-Jun* induction in response to TGF-β (Pertovaara et al., 1989).

It is also consistent with what has been recently reported for the *Smad7* promoter. A report published earlier this year showed, by chromatin immunoprecipitation (ChIP), that SnoN associates with the SBR region of this promoter in the absence of TGF-β but is removed by 15-30 minutes of TGF-β signalling. Following re-synthesis of SnoN, it returns to the Smad7 promoter and is presumably involved in shutting down *Smad7* induction. This might occur once enough Smad7 has been synthesised to shut down the receptors and may be necessary to allow another round of TGF-β signalling once more receptors have become available (Briones-Orta et al., 2006).

In principle, two-step affinity purification as described above but using HA-tagged SnoN instead of Smad3 could be used to identify all factors associated with the *c-Jun* SBR in the absence of TGF-β. This could be expected to lead to a better understanding of the regulatory role of this element in the absence of TGF-β.

5.8 Future perspectives

In summary, this thesis paves the way for more detailed investigation of Smad-transcription factor interactions in fully differentiated cells of the mammalian adult. It describes a purification system that – in some form – should allow identification of novel Smad partners. Subsequent analysis of these factors should then lead to new insights into the TGF-β signalling pathway. In particular, it would be expected to reveal a subset of TGF-β target genes that are regulated through these factors. These genes might be found to show a similar pattern of regulation in response to TGF-β and to be regulated by crosstalk with other pathways in a similar way. Detailed investigation of novel partners would also be expected to yield mechanistic insights into Smad-mediated transcriptional control. The similarities and differences between the interaction of factor X with Smad3 and the Mixer SIM with Smad2, for example, might reveal general principles of Smad-partner interactions. In addition, if factor X associates with known coregulator complexes it would reveal one way in which Smad-mediated gene induction takes place. Finally, some of the identified factors may be subject to genetic or epigenetic changes during the process of tumourigenesis and may regulate a subset of TGF-β gene responses involved in tumour suppression

or tumour promotion. If this were found to be the case, it would constitute a major advance in our understanding of the epigenetic changes that cause cells to switch from a cytostatic response to TGF-β to a metastatic response during the process of tumourigenesis

REFERENCES

Abreu, J. G., Ketpura, N. I., Reversade, B., and De Robertis, E. M. (2002). Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-β. Nat Cell Biol *4*, 599-604.

Akhurst, R. J., and Derynck, R. (2001). TGF-β signaling in cancer - a double-edged sword. Trends Cell Biol *11*, S44-51.

Akiyoshi, S., Inoue, H., Hanai, J., Kusanagi, K., Nemoto, N., Miyazono, K., and Kawabata, M. (1999). c-Ski acts as a transcriptional co-repressor in transforming growth factor-β signaling through interaction with smads. J Biol Chem *274*, 35269-35277.

An, W., Kim, J., and Roeder, R. G. (2004). Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. Cell *117*, 735-748.

Anzano, M. A., Roberts, A. B., Smith, J. M., Sporn, M. B., and De Larco, J. E. (1983). Sarcoma growth factor from conditioned medium of virally transformed cells is composed of both type alpha and type beta transforming growth factors. Proc Natl Acad Sci U S A *80*, 6264-6268.

Ashcroft, G. S. (1999). Bidirectional regulation of macrophage function by TGF-β. Microbes Infect *1*, 1275-1282.

Ashcroft, G. S., Yang, X., Glick, A. B., Weinstein, M., Letterio, J. L., Mizel, D. E., Anzano, M., Greenwell-Wild, T., Wahl, S. M., Deng, C., and Roberts, A. B. (1999). Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. Nat Cell Biol *1*, 260-266.

Bai, R. Y., Koester, C., Ouyang, T., Hahn, S. A., Hammerschmidt, M., Peschel, C., and Duyster, J. (2002). SMIF, a Smad4-interacting protein that functions as a co-activator in TGF-β signalling. Nat Cell Biol *4*, 181-190.

Bakin, A. V., Rinehart, C., Tomlinson, A. K., and Arteaga, C. L. (2002). p38 mitogen-activated protein kinase is required for TGF-β-mediated fibroblastic transdifferentiation and cell migration. J Cell Sci *115*, 3193-3206.

Bakin, A. V., Tomlinson, A. K., Bhowmick, N. A., Moses, H. L., and Arteaga, C. L. (2000). Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration. J Biol Chem *275*, 36803-36810.

Bar-Sagi, D., and Hall, A. (2000). Ras and Rho GTPases: a family reunion. Cell *103*, 227-238.

References

Blobe, G. C., Schiemann, W. P., and Lodish, H. F. (2000). Role of transforming growth factor β in human disease. N Engl J Med *342*, 1350-1358.

Boesen, C. C., Radaev, S., Motyka, S. A., Patamawenu, A., and Sun, P. D. (2002). The 1.1 A crystal structure of human TGF-β type II receptor ligand binding domain. Structure *10*, 913-919.

Bonni, S., Wang, H. R., Causing, C. G., Kavsak, P., Stroschein, S. L., Luo, K., and Wrana, J. L. (2001). TGF-β induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation. Nat Cell Biol *3*, 587-595.

Brenner, C., and Fuks, F. (2006). DNA methyltransferases: facts, clues, mysteries. Curr Top Microbiol Immunol *301*, 45-66.

Briones-Orta, M. A., Sosa-Garrocho, M., Moreno-Alvarez, P., Fonseca-Sanchez, M. A., and Macias-Silva, M. (2006). SnoN co-repressor binds and represses smad7 gene promoter. Biochem Biophys Res Commun *341*, 889-894.

Brodin, G., Ahgren, A., ten Dijke, P., Heldin, C. H., and Heuchel, R. (2000). Efficient TGF-β induction of the Smad7 gene requires cooperation between AP-1, Sp1, and Smad proteins on the mouse Smad7 promoter. J Biol Chem *275*, 29023-29030.

Brown, C. B., Boyer, A. S., Runyan, R. B., and Barnett, J. V. (1999). Requirement of type III TGF-β receptor for endocardial cell transformation in the heart. Science *283*, 2080-2082.

Cano, A., Perez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., and Nieto, M. A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol *2*, 76-83.

Chacko, B. M., Qin, B., Correia, J. J., Lam, S. S., de Caestecker, M. P., and Lin, K. (2001). The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. Nat Struct Biol *8*, 248-253.

Chan, H. M., Shikama, N., and La Thangue, N. B. (2001). Control of gene expression and the cell cycle. Essays Biochem *37*, 87-96.

Chang, H., Brown, C. W., and Matzuk, M. M. (2002). Genetic analysis of the mammalian transforming growth factor-β superfamily. Endocr Rev *23*, 787-823.

References

Chen, A. J., Zhou, G., Juan, T., Colicos, S. M., Cannon, J. P., Cabriera-Hansen, M., Meyer, C. F., Jurecic, R., Copeland, N. G., Gilbert, D. J., *et al.* (2002a). The dual specificity JKAP specifically activates the c-Jun N-terminal kinase pathway. J Biol Chem *277*, 36592-36601.

Chen, C. R., Kang, Y., and Massague, J. (2001). Defective repression of c-myc in breast cancer cells: A loss at the core of the transforming growth factor β growth arrest program. Proc Natl Acad Sci U S A *98*, 992-999.

Chen, C. R., Kang, Y., Siegel, P. M., and Massague, J. (2002b). E2F4/5 and p107 as Smad cofactors linking the TGF-β receptor to c-myc repression. Cell *110*, 19-32.

Chen, D., Zhao, M., and Mundy, G. R. (2004). Bone morphogenetic proteins. Growth Factors *22*, 233-241.

Chen, L. (1999). Combinatorial gene regulation by eukaryotic transcription factors. Curr Opin Struct Biol *9*, 48-55.

Chen, R. H., and Chang, T. Y. (1997). Involvement of caspase family proteases in transforming growth factor-β-induced apoptosis. Cell Growth Differ *8*, 821-827.

Chen, X., Rubock, M. J., and Whitman, M. (1996). A transcriptional partner for MAD proteins in TGF-β signalling. Nature *383*, 691-696.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature *389*, 85-89.

Chen, Y. G., Hata, A., Lo, R. S., Wotton, D., Shi, Y., Pavletich, N., and Massagué, J. (1998). Determinants of specificity in TGF-β signal transduction. Genes Dev *12*, 2144-2152.

Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986). A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol Cell Biol *6*, 4723-4733.

Choo, Y., and Klug, A. (1997). Physical basis of a protein-DNA recognition code. Curr Opin Struct Biol *7*, 117-125.

Chu, B., Soncin, F., Price, B. D., Stevenson, M. A., and Calderwood, S. K. (1996). Sequential phosphorylation by mitogen-activated protein kinase and glycogen synthase kinase 3 represses transcriptional activation by heat shock factor-1. J Biol Chem *271*, 30847-30857.

References

Chu, G. C., Dunn, N. R., Anderson, D. C., Oxburgh, L., and Robertson, E. J. (2004). Differential requirements for Smad4 in TGF-β-dependent patterning of the early mouse embryo. Development *131*, 3501-3512.

Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T., and Conaway, J. W. (2005). The mammalian Mediator complex and its role in transcriptional regulation. Trends Biochem Sci *30*, 250-255.

Conery, A. R., Cao, Y., Thompson, E. A., Townsend, C. M., Jr., Ko, T. C., and Luo, K. (2004). Akt interacts directly with Smad3 to regulate the sensitivity to TGF-β induced apoptosis. Nat Cell Biol *6*, 366-372.

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., and Piccolo, S. (2003). Links between tumor suppressors: p53 is required for TGF-β gene responses by cooperating with Smads. Cell *113*, 301-314.

Courey, A. J. (2001). Cooperativity in transcriptional control. Curr Biol *11*, R250-252.

Crawford, S. E., Stellmach, V., Murphy-Ullrich, J. E., Ribeiro, S. M., Lawler, J., Hynes, R. O., Boivin, G. P., and Bouck, N. (1998). Thrombospondin-1 is a major activator of TGF-β1 in vivo. Cell *93*, 1159-1170.

Crease, D. J., Dyson, S., and Gurdon, J. B. (1998). Cooperation between the activin and Wnt pathways in the spatial control of organizer gene expression. Proc Natl Acad Sci U S A *95*, 4398-4403.

Cui, W., Fowlis, D. J., Bryson, S., Duffie, E., Ireland, H., Balmain, A., and Akhurst, R. J. (1996). TGFβ1 inhibits the formation of benign skin tumors, but enhances progression to invasive spindle carcinomas in transgenic mice. Cell *86*, 531-542.

Czarniecki, C. W., Chiu, H. H., Wong, G. H., McCabe, S. M., and Palladino, M. A. (1988). Transforming growth factor-β 1 modulates the expression of class II histocompatibility antigens on human cells. J Immunol *140*, 4217-4223.

De Cesare, D., Fimia, G. M., and Sassone-Corsi, P. (1999). Signaling routes to CREM and CREB: plasticity in transcriptional activation. Trends Biochem Sci *24*, 281-285.

Denissova, N. G., Pouponnot, C., Long, J., He, D., and Liu, F. (2000). Transforming growth factor beta -inducible independent binding of SMAD to the Smad7 promoter. Proc Natl Acad Sci U S A *97*, 6397-6402.

Dennler, S., Prunier, C., Ferrand, N., Gauthier, J. M., and Atfi, A. (2000). c-Jun inhibits transforming growth factor beta-mediated transcription by repressing Smad3 transcriptional activity. J Biol Chem *275*, 28858-28865.

Derynck, R., Akhurst, R. J., and Balmain, A. (2001). TGF-β signaling in tumor suppression and cancer progression. Nat Genet *29*, 117-129.

Derynck, R., and Zhang, Y. E. (2003). Smad-dependent and Smad-independent pathways in TGF-β family signalling. Nature *425*, 577-584.

Di Guglielmo, G. M., Le Roy, C., Goodfellow, A. F., and Wrana, J. L. (2003). Distinct endocytic pathways regulate TGF-β receptor signalling and turnover. Nat Cell Biol *5*, 410-421.

Dickson, M. C., Martin, J. S., Cousins, F. M., Kulkarni, A. B., Karlsson, S., and Akhurst, R. J. (1995). Defective haematopoiesis and vasculogenesis in transforming growth factor-β1 knock out mice. Development *121*, 1845-1854.

Dimova, D. K., and Dyson, N. J. (2005). The E2F transcriptional network: old acquaintances with new faces. Oncogene *24*, 2810-2826.

Docagne, F., Gabriel, C., Lebeurrier, N., Lesne, S., Hommet, Y., Plawinski, L., Mackenzie, E. T., and Vivien, D. (2004). Sp1 and Smad transcription factors co-operate to mediate TGF-β-dependent activation of amyloid-β precursor protein gene transcription. Biochem J *383*, 393-399.

Doyon, Y., Selleck, W., Lane, W. S., Tan, S., and Cote, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol *24*, 1884-1896.

Duivenvoorden, W. C., Hirte, H. W., and Singh, G. (1999). Transforming growth factor beta1 acts as an inducer of matrix metalloproteinase expression and activity in human bone-metastasizing cancer cells. Clin Exp Metastasis *17*, 27-34.

Dumont, J. E., Pecasse, F., and Maenhaut, C. (2001). Crosstalk and specificity in signalling. Are we crosstalking ourselves into general confusion? Cell Signal *13*, 457-463.

Dunn, N. R., Koonce, C. H., Anderson, D. C., Islam, A., Bikoff, E. K., and Robertson, E. J. (2005). Mice exclusively expressing the short isoform of Smad2 develop normally and are viable and fertile. Genes Dev *19*, 152-163.

References

Dynan, W. S., and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell *32*, 669-680.

Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T., and Miyazono, K. (2001). Smurf1 interacts with transforming growth factor-β type I receptor through Smad7 and induces receptor degradation. J Biol Chem *276*, 12477-12480.

Edlund, S., Landstrom, M., Heldin, C. H., and Aspenstrom, P. (2002). Transforming growth factor-β-induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA. Mol Biol Cell *13*, 902-914.

Engel, M. E., Datta, P. K., and Moses, H. L. (1998). RhoB is stabilized by transforming growth factor β and antagonizes transcriptional activation. J Biol Chem *273*, 9921-9926.

Felici, A., Wurthner, J. U., Parks, W. T., Giam, L. R., Reiss, M., Karpova, T. S., McNally, J. G., and Roberts, A. B. (2003). TLP, a novel modulator of TGF-β signaling, has opposite effects on Smad2- and Smad3-dependent signaling. EMBO J *22*, 4465-4477.

Feng, X. H., Lin, X., and Derynck, R. (2000). Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(Ink4B) transcription in response to TGF-β. EMBO J *19*, 5178-5193.

Feng, X. H., Zhang, Y., Wu, R. Y., and Derynck, R. (1998). The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-β-induced transcriptional activation. Genes Dev *12*, 2153-2163.

Fink, S. P., Mikkola, D., Willson, J. K., and Markowitz, S. (2003). TGF-β-induced nuclear localization of Smad2 and Smad3 in Smad4 null cancer cell lines. Oncogene *22*, 1317-1323.

Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr Opin Genet Dev *14*, 165-173.

Gajiwala, K. S., and Burley, S. K. (2000). Winged helix proteins. Curr Opin Struct Biol *10*, 110-116.

Germain, S., Howell, M., Esslemont, G. M., and Hill, C. S. (2000). Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev *14*, 435-451.

Ghosh, A. K., Bhattacharyya, S., and Varga, J. (2004). The tumor suppressor p53 abrogates Smad-dependent collagen gene induction in mesenchymal cells. J Biol Chem *279*, 47455-47463.

Giese, K., Amsterdam, A., and Grosschedl, R. (1991). DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes Dev *5*, 2567-2578.

Gilboa, L., Wells, R. G., Lodish, H. F., and Henis, Y. I. (1998). Oligomeric structure of type I and type II transforming growth factor beta receptors: homodimers form in the ER and persist at the plasma membrane. J Cell Biol *140*, 767-777.

Giniger, E., and Ptashne, M. (1987). Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. Nature *330*, 670-672.

Gonzalez, G. A., and Montminy, M. R. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell *59*, 675-680.

Gorelik, L., and Flavell, R. A. (2002). Transforming growth factor-β in T-cell biology. Nat Rev Immunol *2*, 46-53.

Gotzmann, J., Fischer, A. N., Zojer, M., Mikula, M., Proell, V., Huber, H., Jechlinger, M., Waerner, T., Weith, A., Beug, H., and Mikulits, W. (2006). A crucial function of PDGF in TGF-β-mediated cancer progression of hepatocytes. Oncogene.

Goumans, M. J., Valdimarsdottir, G., Itoh, S., Lebrin, F., Larsson, J., Mummery, C., Karlsson, S., and ten Dijke, P. (2003). Activin receptor-like kinase (ALK)1 is an antagonistic mediator of lateral TGF-β/ALK5 signaling. Mol Cell *12*, 817-828.

Grant, P. A., and Berger, S. L. (1999). Histone acetyltransferase complexes. Semin Cell Dev Biol *10*, 169-177.

Greenwald, J., Fischer, W. H., Vale, W. W., and Choe, S. (1999). Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. Nat Struct Biol *6*, 18-22.

Gu, W., Malik, S., Ito, M., Yuan, C. X., Fondell, J. D., Zhang, X., Martinez, E., Qin, J., and Roeder, R. G. (1999). A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. Mol Cell *3*, 97-108.

Guaita, S., Puig, I., Franci, C., Garrido, M., Dominguez, D., Batlle, E., Sancho, E., Dedhar, S., De Herreros, A. G., and Baulida, J. (2002). Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. J Biol Chem *277*, 39209-39216.

References

Hall, W. W., and Fujii, M. (2005). Deregulation of cell-signaling pathways in HTLV-1 infection. Oncogene *24*, 5965-5975.

Hannon, G. J., and Beach, D. (1994). p15INK4B is a potential effector of TGF-β-induced cell cycle arrest. Nature *371*, 257-261.

Harlow, E. and Lane, D. (1988). Antibodies, a Laboratory Manual. Cold Spring Harbour Laboratory Press.

Hart, P. J., Deep, S., Taylor, A. B., Shu, Z., Hinck, C. S., and Hinck, A. P. (2002). Crystal structure of the human TβRII ectodomain-TGF-β3 complex. Nat Struct Biol *9*, 203-208.

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., and Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. Cell *111*, 369-379.

Hata, A., Lo, R. S., Wotton, D., Lagna, G., and Massagué, J. (1997). Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. Nature *388*, 82-87.

Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y.-Y., Grinnell, B. W., Richardson, M. A., Topper, J. N., Gimbrone, M. A. J., Wrana, J., and Falb, D. (1997). The MAD-Related Protein Smad7 Associates with the TGFβ Receptor and Functions as an Antagonist of TGFβ Signaling. Cell *89*, 1165-1173.

Helmreich, J. M. (2001). The Biochemistry of Cell Signalling, 1st. ed. Oxford University Press.

Hess, J., Angel, P., and Schorpp-Kistner, M. (2004). AP-1 subunits: quarrel and harmony among siblings. J Cell Sci *117*, 5965-5973.

Heyer, J., Escalante-Alcalde, D., Lia, M., Boettinger, E., Edelmann, W., Stewart, C. L., and Kucherlapati, R. (1999). Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. Proc Natl Acad Sci U S A *96*, 12595-12600.

Hill, C. S. (2001). TGF-β signalling in early *Xenopus* development. Cur Opin Genet Dev *11*, 533-540.

Hoffmann, I., Draetta, G., and Karsenti, E. (1994). Activation of the phosphatase activity of human cdc25A by a cdk2-cyclin E dependent phosphorylation at the G1/S transition. EMBO J *13*, 4302-4310.

Hojo, M., Morimoto, T., Maluccio, M., Asano, T., Morimoto, K., Lagman, M., Shimbo, T., and Suthanthiran, M. (1999). Cyclosporine induces cancer progression by a cell-autonomous mechanism. Nature *397*, 530-534.

Hoodless, P. A., Pye, M., Chazaud, C., Labbe, E., Attisano, L., Rossant, J., and Wrana, J. L. (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. Genes Dev *15*, 1257-1271.

Horvai, A. E., Xu, L., Korzus, E., Brard, G., Kalafus, D., Mullen, T. M., Rose, D. W., Rosenfeld, M. G., and Glass, C. K. (1997). Nuclear integration of JAK/STAT and Ras/AP-1 signaling by CBP and p300. Proc Natl Acad Sci U S A *94*, 1074-1079.

Howell, M., Inman, G. J., and Hill, C. S. (2002). A novel *Xenopus* Smad-interacting forkhead transcription factor (XFast-3) cooperates with XFast-1 in regulating gastrulation movements. Development *129*, 2823-2834.

Howell, M., Itoh, F., Pierreux, C. E., Valgeirsdottir, S., Itoh, S., ten Dijke, P., and Hill, C. S. (1999). *Xenopus* Smad4β is the co-Smad component of developmentally-regulated transcription factor complexes responsible for induction of early mesodermal genes. Dev Biol *214*, 354-369.

Howell, M., Mohun, T. J., and Hill, C. S. (2001). Xenopus Smad3 is specifically expressed in the chordoneural hinge, notochord and in the endocardium of the developing heart. Mech Dev *104*, 147-150.

Huse, M., Chen, Y. G., Massague, J., and Kuriyan, J. (1999). Crystal structure of the cytoplasmic domain of the type I TGF-β receptor in complex with FKBP12. Cell *96*, 425-436.

Huse, M., Muir, T. W., Xu, L., Chen, Y. G., Kuriyan, J., and Massague, J. (2001). The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. Mol Cell *8*, 671-682.

Iavarone, A., and Massagué, J. (1997). Repression of the CDK activator Cdc25A and cell-cycle arrest by cytokine TGF-β in cells lacking the CDK inhibitor p15. Nature *387*, 417-422.

Iizuka, M., and Smith, M. M. (2003). Functional consequences of histone modifications. Curr Opin Genet Dev *13*, 154-160.

Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M., and Miyazono, K. (1997). Smad6 inhibits signalling by the TGF-β superfamily. Nature *389*, 622-626.

Indra, A. K., Mohan, W. S., 2nd, Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., and Tora, L. (2005). TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis. Dev Biol *285*, 28-37.

Inman, G. J., and Hill, C. S. (2002). Stoichiometry of active smad-transcription factor complexes on DNA. J Biol Chem *277*, 51008-51016.

Inman, G. J., Nicolás, F. J., and Hill, C. S. (2002). Nucleocytoplasmic shuttling of Smads 2, 3 and 4 permits sensing of TGF-β receptor activity. Mol Cell *10*, 283-294.

Ishida, W., Hamamoto, T., Kusanagi, K., Yagi, K., Kawabata, M., Takehara, K., Sampath, T. K., Kato, M., and Miyazono, K. (2000). Smad6 is a Smad1/5-induced smad inhibitor. Characterization of bone morphogenetic protein-responsive element in the mouse Smad6 promoter. J Biol Chem *275*, 6075-6079.

Itoh, S., Landstrom, M., Hermansson, A., Itoh, F., Heldin, C. H., Heldin, N. E., and ten Dijke, P. (1998). Transforming growth factor β1 induces nuclear export of inhibitory Smad7. J Biol Chem *273*, 29195-29201.

Jackson, M. B. (1999). Ligand-gated channel: postsynaptic receptors and drug targets. Adv Neurol *79*, 511-524.

Jacobs, S. A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jinnin, M., Ihn, H., Asano, Y., Yamane, K., Trojanowska, M., and Tamaki, K. (2004). Tenascin-C upregulation by transforming growth factor-β in human dermal fibroblasts involves Smad3, Sp1, and Ets1. Oncogene *23*, 1656-1667.

Johnson, K., Kirkpatrick, H., Comer, A., Hoffmann, F. M., and Laughon, A. (1999). Interaction of Smad complexes with tripartite DNA-binding sites. J Biol Chem *274*, 20709-20716.

Kaartinen, V., Voncken, J. W., Shuler, C., Warburton, D., Bu, D., Heisterkamp, N., and Groffen, J. (1995). Abnormal lung development and cleft palate in mice lacking TGF-β3 indicates defects of epithelial-mesenchymal interaction. Nat Genet *11*, 415-421.

Kadesch, T. (2004). Notch signaling: the demise of elegant simplicity. Curr Opin Genet Dev *14*, 506-512.

Kadonaga, J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. Cell *116*, 247-257.

Kaestner, K. H., Knochel, W., and Martinez, D. E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev *14*, 142-146.

Kang, J. S., Alliston, T., Delston, R., and Derynck, R. (2005). Repression of Runx2 function by TGF-β through recruitment of class II histone deacetylases by Smad3. EMBO J *24*, 2543-2555.

Kang, Y., Chen, C. R., and Massague, J. (2003). A self-enabling TGF-β response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Mol Cell *11*, 915-926.

Kato, Y., Habas, R., Katsuyama, Y., Naar, A. M., and He, X. (2002). A component of the ARC/Mediator complex required for TGF-β/Nodal signalling. Nature *418*, 641-646.

Kaufmann, E., and Knochel, W. (1996). Five years on the wings of fork head. Mech Dev *57*, 3-20.

Kavsak, P., Rasmussen, R. K., Causing, C. G., Bonni, S., Zhu, H., Thomsen, G. H., and Wrana, J. L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF β receptor for degradation. Mol Cell *6*, 1365-1375.

Kehrl, J. H., Roberts, A. B., Wakefield, L. M., Jakowlew, S., Sporn, M. B., and Fauci, A. S. (1986a). Transforming growth factor beta is an important immunomodulatory protein for human B lymphocytes. J Immunol *137*, 3855-3860.

Kehrl, J. H., Wakefield, L. M., Roberts, A. B., Jakowlew, S., Alvarez-Mon, M., Derynck, R., Sporn, M. B., and Fauci, A. S. (1986b). Production of transforming growth factor β by human T lymphocytes and its potential role in the regulation of T cell growth. J Exp Med *163*, 1037-1050.

Kim, J., Johnson, K., Chen, H. J., Carroll, S., and Laughon, A. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. Nature *388*, 304-308.

Kim, J. H., Lane, W. S., and Reinberg, D. (2002). Human Elongator facilitates RNA polymerase II transcription through chromatin. Proc Natl Acad Sci U S A *99*, 1241-1246.

Kim, R. H., Wang, D., Tsang, M., Martin, J., Huff, C., de Caestecker, M. P., Parks, W. T., Meng, X., Lechleider, R. J., Wang, T., and Roberts, A. B. (2000). A novel smad nuclear interacting protein, SNIP1, suppresses p300-dependent TGF-β signal transduction. Genes Dev *14*, 1605-1616.

Kirsch, T., Sebald, W., and Dreyer, M. K. (2000). Crystal structure of the BMP-2-BMPRIA ectodomain complex. Nat Struct Biol *7*, 492-496.

Kon, A., Vindevoghel, L., Kouba, D. J., Fujimura, Y., Uitto, J., and Mauviel, A. (1999). Cooperation between SMAD and NF-kB in growth factor regulated type VII collagen gene expression. Oncogene *18*, 1837-1844.

Kondo, M., Cubillo, E., Tobiume, K., Shirakihara, T., Fukuda, N., Suzuki, H., Shimizu, K., Takehara, K., Cano, A., Saitoh, M., and Miyazono, K. (2004). A role for Id in the regulation of TGF-β-induced epithelial-mesenchymal transdifferentiation. Cell Death Differ *11*, 1092-1101.

Korchynskyi, O., and ten Dijke, P. (2002). Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. J Biol Chem *277*, 4883-4891.

Kordula, T., Guttgemann, I., Rose-John, S., Roeb, E., Osthues, A., Tschesche, H., Koj, A., Heinrich, P. C., and Graeve, L. (1992). Synthesis of tissue inhibitor of metalloproteinase-1 (TIMP-1) in human hepatoma cells (HepG2). Up-regulation by interleukin-6 and transforming growth factor beta 1. FEBS Lett *313*, 143-147.

Kretzschmar, M., Doody, J., Timokhina, I., and Massagué, J. (1999). A mechanism of repression of TGFβ/Smad signaling by oncogenic Ras. Genes Dev *13*, 804-816.

Kulkarni, A. B., Huh, C. G., Becker, D., Geiser, A., Lyght, M., Flanders, K. C., Roberts, A. B., Sporn, M. B., Ward, J. M., and Karlsson, S. (1993). Transforming growth factor β 1 null mutation in mice causes excessive inflammatory response and early death. Proc Natl Acad Sci U S A *90*, 770-774.

Kunwar, P. S., Zimmerman, S., Bennett, J. T., Chen, Y., Whitman, M., and Schier, A. F. (2003). Mixer/Bon and FoxH1/Sur have overlapping and divergent roles in Nodal signaling and mesendoderm induction. Development *130*, 5589-5599.

Kurisaki, A., Kose, S., Yoneda, Y., Heldin, C. H., and Moustakas, A. (2001). Transforming growth factor-β induces nuclear import of Smad3 in an importin-β1 and Ran-dependent manner. Mol Biol Cell *12*, 1079-1091.

Kurisaki, A., Kurisaki, K., Kowanetz, M., Sugino, H., Yoneda, Y., Heldin, C. H., and Moustakas, A. (2006). The mechanism of nuclear export of Smad3 involves exportin 4 and Ran. Mol Cell Biol *26*, 1318-1332.

Labbe, E., Letamendia, A., and Attisano, L. (2000). Association of Smads with lymphoid enhancer binding factor 1/T cell- specific factor mediates cooperative signaling by the transforming growth factor-β and wnt pathways. Proc Natl Acad Sci U S A *97*, 8358-8363.

Labbe, E., Silvestri, C., Hoodless, P. A., Wrana, J. L., and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGF β-dependent transcription through the forkhead DNA-binding protein FAST2. Mol Cell *2*, 109-120.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laiho, M., DeCaprio, J. A., Ludlow, J. W., Livingston, D. M., and Massagué, J. (1990). Growth inhibition by TGF-β linked to suppression of retinoblastoma protein phosphorylation. Cell *62*, 175-185.

Laiho, M., Saksela, O., Andreasen, P. A., and Keski-Oja, J. (1986). Enhanced production and extracellular deposition of the endothelial- type plasminogen activator inhibitor in cultured human lung fibroblasts by transforming growth factor-β. J Cell Biol *103*, 2403-2410.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lasorella, A., Noseda, M., Beyna, M., Yokota, Y., and Iavarone, A. (2000). Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins. Nature *407*, 592-598.

Latchmann, D. (2002). Gene Regulation, A Eukaryotic Perspective, 4th ed. Nelson Thornes.

Lebman, D. A., and Edmiston, J. S. (1999). The role of TGF-β in growth, differentiation, and maturation of B lymphocytes. Microbes Infect *1*, 1297-1304.

Lebrin, F., Goumans, M. J., Jonker, L., Carvalho, R. L., Valdimarsdottir, G., Thorikay, M., Mummery, C., Arthur, H. M., and ten Dijke, P. (2004). Endoglin promotes endothelial cell proliferation and TGF-β/ALK1 signal transduction. EMBO J *23*, 4018-4028.

Lee, M., and Goodbourn, S. (2001). Signalling from the cell surface to the nucleus. Essays Biochem *37*, 71-85.

Lee, M. H., and Yang, H. Y. (2001). Negative regulators of cyclin-dependent kinases and their roles in cancers. Cell Mol Life Sci *58*, 1907-1922.

Leevers, S. J., Vanhaesebroeck, B., and Waterfield, M. D. (1999). Signalling through phosphoinositide 3-kinases: the lipids take centre stage. Curr Opin Cell Biol *11*, 219-225.

Lehmann, K., Janda, E., Pierreux, C. E., Rytomaa, M., Schulze, A., McMahon, M., Hill, C. S., Beug, H., and Downward, J. (2000). Raf induces TGFβ production while blocking its apoptotic but not invasive responses: a mechanism leading to increased malignancy in epithelial cells. Genes Dev *14*, 2610-2622.

Levy, L., and Hill, C. S. (2005). Smad4 dependency defines two classes of transforming growth factor β (TGF-β) target genes and distinguishes TGF-β-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. Mol Cell Biol *25*, 8108-8125.

Levy, L., and Hill, C. S. (2006). Alterations in components of the TGF-β superfamily signaling pathways in human cancer. Cytokine Growth Factor Rev *17*, 41-58.

Li, L., Hu, J. S., and Olson, E. N. (1990). Different members of the jun proto-oncogene family exhibit distinct patterns of expression in response to type beta transforming growth factor. J Biol Chem *265*, 1556-1562.

Liberati, N. T., Datto, M. B., Frederick, J. P., Shen, X., Wong, C., Rougier-Chapman, E. M., and Wang, X. F. (1999). Smads bind directly to the Jun family of AP-1 transcription factors. Proc Natl Acad Sci U S A *96*, 4844-4849.

Liu, D., Black, B. L., and Derynck, R. (2001). TGF-β inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. Genes Dev *15*, 2950-2966.

Liu, D., Kang, J. S., and Derynck, R. (2004). TGF-β-activated Smad3 represses MEF2-dependent transcription in myogenic differentiation. EMBO J *23*, 1557-1566.

Lo, R. S., Chen, Y. G., Shi, Y., Pavletich, N. P., and Massague, J. (1998). The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-β receptors. EMBO J *17*, 996-1005.

Lo, R. S., Wotton, D., and Massague, J. (2001). Epidermal growth factor signaling via Ras controls the Smad transcriptional co-repressor TGIF. EMBO J *20*, 128-136.

Lopez-Rovira, T., Chalaux, E., Massague, J., Rosa, J. L., and Ventura, F. (2002). Direct binding of Smad1 and Smad4 to two distinct motifs mediates bone morphogenetic protein-specific transcriptional activation of Id1 gene. J Biol Chem *277*, 3176-3185.

Luo, K. (2004). Ski and SnoN: negative regulators of TGF-β signaling. Curr Opin Genet Dev *14*, 65-70.

Luo, K., and Lodish, H. F. (1996). Signaling by chimeric erythropoietin-TGF-β receptors: homodimerization of the cytoplasmic domain of the type I TGF-β receptor and heterodimerization with the type II receptor are both required for intracellular signal transduction. EMBO J *15*, 4485-4496.

Luo, K., and Lodish, H. F. (1997). Positive and negative regulation of type II TGF-β receptor signal transduction by autophosphorylation on multiple serine residues. EMBO J *16*, 1970-1981.

Luo, K., Stroschein, S. L., Wang, W., Chen, D., Martens, E., Zhou, S., and Zhou, Q. (1999). The Ski oncoprotein interacts with the Smad proteins to repress TGF-β signaling. Genes Dev *13*, 2196-2206.

Lyons, R. M., Gentry, L. E., Purchio, A. F., and Moses, H. L. (1990). Mechanism of activation of latent recombinant transforming growth factor β1 by plasmin. J Cell Biol *110*, 1361-1367.

Macías-Silva, M., Abdollah, S., Hoodless, P. A., Pirone, R., Attisano, L., and Wrana, J. L. (1996). MADR2 is a substrate of the TGFβ receptor and its phosphorylation is required for nuclear accumulation and signaling. Cell *87*, 1215-1224.

Madri, J. A., Pratt, B. M., and Tucker, A. M. (1988). Phenotypic modulation of endothelial cells by transforming growth factor-β depends upon the composition and organization of the extracellular matrix. J Cell Biol *106*, 1375-1384.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Marais, R., Wynne, J., and Treisman, R. (1993). The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. Cell *73*, 381-393.

Marinissen, M. J., and Gutkind, J. S. (2001). G-protein-coupled receptors and signaling networks: emerging paradigms. Trends Pharmacol Sci *22*, 368-376.

Martin, J. S., Dickson, M. C., Cousins, F. M., Kulkarni, A. B., Karlsson, S., and Akhurst, R. J. (1995). Analysis of homozygous TGF-β1 null mouse embryos demonstrates defects in yolk sac vasculogenesis and hematopoiesis. Ann N Y Acad Sci *752*, 300-308.

Massague, J. (1999). Wounding Smad. Nat Cell Biol *1*, E117-119.

Massague, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev *19*, 2783-2810.

Massagué, J., and Wotton, D. (2000). Transcriptional control by the TGF-β/Smad signaling system. EMBO J *19*, 1745-1754.

Matsuura, I., Denissova, N. G., Wang, G., He, D., Long, J., and Liu, F. (2004). Cyclin-dependent kinases regulate the antiproliferative function of Smads. Nature *430*, 226-231.

Mayr, B., and Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. Nat Rev Mol Cell Biol *2*, 599-609.

McKenna, N. J., and O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. Cell *108*, 465-474.

Melhuish, T. A., Gallo, C. M., and Wotton, D. (2001). TGIF2 interacts with histone deacetylase 1 and represses transcription. J Biol Chem *276*, 32109-32114.

Melhuish, T. A., and Wotton, D. (2000). The interaction of the carboxyl terminus-binding protein with the Smad corepressor TGIF is disrupted by a holoprosencephaly mutation in TGIF. J Biol Chem *275*, 39762-39766.

Morisato, D., and Anderson, K. V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo. Annu Rev Genet *29*, 371-399.

Munger, J. S., Huang, X., Kawakatsu, H., Griffiths, M. J., Dalton, S. L., Wu, J., Pittet, J. F., Kaminski, N., Garat, C., Matthay, M. A., *et al.* (1999). The integrin αvβ6 binds and activates latent TGF-β1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell *96*, 319-328.

Nakajima, T., Uchida, C., Anderson, S. F., Parvin, J. D., and Montminy, M. (1997). Analysis of a cAMP-responsive activator reveals a two-component mechanism for transcriptional induction via signal-dependent factors. Genes Dev *11*, 738-747.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J. L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N. E., Heldin, C. H., and ten Dijke, P. (1997). Identification of Smad7, a TGFβ-inducible antagonist of TGF-β signalling. Nature *389*, 631-635.

Narlikar, G. J., Fan, H. Y., and Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nicolas, F. J., De Bosscher, K., Schmierer, B., and Hill, C. S. (2004). Analysis of Smad nucleocytoplasmic shuttling in living cells. J Cell Sci *117*, 4113-4125.

Nieto, M. A. (2002). The snail superfamily of zinc-finger transcription factors. Nat Rev Mol Cell Biol *3*, 155-166.

Nishihara, A., Hanai, J., Imamura, T., Miyazono, K., and Kawabata, M. (1999). E1A inhibits transforming growth factor-β signaling through binding to Smad proteins. J Biol Chem *274*, 28716-28723.

Nishita, M., Hashimoto, M. K., Ogata, S., Laurent, M. N., Ueno, N., Shibuya, H., and Cho, K. W. (2000). Interaction between Wnt and TGF-β signalling pathways during formation of Spemann's organizer. Nature *403*, 781-785.

Oft, M., Heider, K. H., and Beug, H. (1998). TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis. Curr Biol *8*, 1243-1252.

Oh, S. P., Seki, T., Goss, K. A., Imamura, T., Yi, Y., Donahoe, P. K., Li, L., Miyazono, K., ten Dijke, P., Kim, S., and Li, E. (2000). Activin receptor-like kinase 1 modulates transforming growth factor-β1 signaling in the regulation of angiogenesis. Proc Natl Acad Sci U S A *97*, 2626-2631.

Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105-116.

Ozdamar, B., Bose, R., Barrios-Rodiles, M., Wang, H. R., Zhang, Y., and Wrana, J. L. (2005). Regulation of the polarity protein Par6 by TGF-β receptors controls epithelial cell plasticity. Science *307*, 1603-1609.

Pabo, C. O., and Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem *61*, 1053-1095.

Pan, J., Clayton, M., and Feitelson, M. A. (2004). Hepatitis B virus X antigen promotes transforming growth factor-β1 (TGF-β1) activity by up-regulation of TGF-β1 and down-regulation of alpha2-macroglobulin. J Gen Virol *85*, 275-282.

Papamichos-Chronakis, M., Conlan, R. S., Gounalaki, N., Copf, T., and Tzamarias, D. (2000). Hrs1/Med3 is a Cyc8-Tup1 corepressor target in the RNA polymerase II holoenzyme. J Biol Chem *275*, 8397-8403.

Pardali, K., Kurisaki, A., Moren, A., ten Dijke, P., Kardassis, D., and Moustakas, A. (2000). Role of Smad proteins and transcription factor Sp1 in p21(Waf1/Cip1) regulation by transforming growth factor-β. J Biol Chem *275*, 29244-29256.

Pardoux, C., and Derynck, R. (2004). JNK regulates expression and autocrine signaling of TGF-β1. Mol Cell *15*, 170-171.

Paulson, Q. X., McArthur, M. J., and Johnson, D. G. (2006). E2F3a stimulates proliferation, p53-independent apoptosis and carcinogenesis in a transgenic mouse model. Cell Cycle *5*, 184-190.

Pelton, R. W., Saxena, B., Jones, M., Moses, H. L., and Gold, L. I. (1991). Immunohistochemical localization of TGF-β1, TGF-β2, and TGF-β3 in the mouse embryo: expression patterns suggest multiple roles during embryonic development. J Cell Biol *115*, 1091-1105.

Perlman, R., Schiemann, W. P., Brooks, M. W., Lodish, H. F., and Weinberg, R. A. (2001). TGF-β-induced apoptosis is mediated by the adapter protein Daxx that facilitates JNK activation. Nat Cell Biol *3*, 708-714.

Pertovaara, L., Kaipainen, A., Mustonen, T., Orpana, A., Ferrara, N., Saksela, O., and Alitalo, K. (1994). Vascular endothelial growth factor is induced in response to transforming growth factor-β in fibroblastic and epithelial cells. J Biol Chem *269*, 6271-6274.

Pertovaara, L., Sistonen, L., Bos, T. J., Vogt, P. K., Keski-Oja, J., and Alitalo, K. (1989). Enhanced jun gene expression is an early genomic response to transforming growth factor β stimulation. Mol Cell Biol *9*, 1255-1262.

Pessah, M., Marais, J., Prunier, C., Ferrand, N., Lallemand, F., Mauviel, A., and Atfi, A. (2002). c-Jun associates with the oncoprotein Ski and suppresses Smad2 transcriptional activity. J Biol Chem *277*, 29094-29100.

Pessah, M., Prunier, C., Marais, J., Ferrand, N., Mazars, A., Lallemand, F., Gauthier, J. M., and Atfi, A. (2001). c-Jun interacts with the corepressor TG-interacting factor (TGIF) to suppress Smad2 transcriptional activity. Proc Natl Acad Sci U S A *98*, 6198-6203.

Peterson, C. L., and Laniel, M. A. (2004). Histones and histone modifications. Curr Biol *14*, R546-551.
Petritsch, C., Beug, H., Balmain, A., and Oft, M. (2000). TGF-β inhibits p70 S6 kinase via protein phosphatase 2A to induce G(1) arrest. Genes Dev *14*, 3093-3101.

Pierreux, C. E., Nicolás, F. J., and Hill, C. S. (2000). Transforming growth factor β-independent shuttling of Smad4 between the cytoplasm and nucleus. Mol Cell Biol *20*, 9041-9054.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res *18*, 6197-6204.

Portella, G., Cumming, S. A., Liddell, J., Cui, W., Ireland, H., Akhurst, R. J., and Balmain, A. (1998). Transforming growth factor β is essential for spindle cell conversion of mouse skin carcinoma in vivo: implications for tumor invasion. Cell Growth Differ *9*, 393-404.

Pouponnot, C., Jayaraman, L., and Massague, J. (1998). Physical and functional interaction of SMADs and p300/CBP. J Biol Chem *273*, 22865-22868.

Preobrazhenska, O., Yakymovych, M., Kanamoto, T., Yakymovych, I., Stoika, R., Heldin, C. H., and Souchelnytskyi, S. (2002). BRCA2 and Smad3 synergize in regulation of gene transcription. Oncogene *21*, 5660-5664.

Proetzel, G., Pawlowski, S. A., Wiles, M. V., Yin, M., Boivin, G. P., Howles, P. N., Ding, J., Ferguson, M. W., and Doetschman, T. (1995). Transforming growth factor-β3 is required for secondary palate fusion. Nat Genet *11*, 409-414.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683-689.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Qing, J., Zhang, Y., and Derynck, R. (2000). Structural and functional characterization of the transforming growth factor-β-induced Smad3/c-Jun transcriptional cooperativity. J Biol Chem *275*, 38802-38812.

Randall, R. A., Germain, S., Inman, G. J., Bates, P. A., and Hill, C. S. (2002). Different Smad2 partners bind a common hydrophobic pocket in Smad2 via a defined proline-rich motif. EMBO J *21*, 145-156.

Randall, R. A., Howell, M., Page, C. S., Daly, A., Bates, P. A., and Hill, C. S. (2004). Recognition of phosphorylated-Smad2-containing complexes by a novel Smad interaction motif. Mol Cell Biol *24*, 1106-1121.

Reinhold, D., Wrenger, S., Kahne, T., and Ansorge, S. (1999). HIV-1 Tat: immunosuppression via TGF-β1 induction. Immunol Today *20*, 384-385.

Reynisdottir, I., Polyak, K., Iavarone, A., and Massagué, J. (1995). Kip/Cip and Ink4 Cdk inhibitors cooperate to induce cell cycle arrest in response to TGF-β. Genes Dev *9*, 1831-1845.

Rodriguez, C., Chen, F., Weinberg, R. A., and Lodish, H. F. (1995). Cooperative binding of transforming growth factor (TGF)-β2 to the types I and II TGF-β receptors. J Biol Chem *270*, 15919-15922.

Rodriguez, C., Huang, L. J., Son, J. K., McKee, A., Xiao, Z., and Lodish, H. F. (2001). Functional cloning of the proto-oncogene brain factor-1 (BF-1) as a Smad-binding antagonist of transforming growth factor-β signaling. J Biol Chem *276*, 30224-30230.

Safe, S., and Abdelrahim, M. (2005). Sp transcription factor family and its role in cancer. Eur J Cancer *41*, 2438-2448.

Saijoh, Y., Adachi, H., Sakuma, R., Yeo, C. Y., Yashiro, K., Watanabe, M., Hashiguchi, H., Mochida, K., Ohishi, S., Kawabata, M., *et al.* (2000). Left-right asymmetric expression of lefty2 and nodal is induced by a signaling pathway that includes the transcription factor FAST2. Mol Cell *5*, 35-47.

Saltzman, A., Munro, R., Searfoss, G., Franks, C., Jaye, M., and Ivashchenko, Y. (1998). Transforming growth factor-β-mediated apoptosis in the Ramos B-lymphoma cell line is accompanied by caspase activation and Bcl-XL downregulation. Exp Cell Res *242*, 244-254.

Sanford, L. P., Ormsby, I., Gittenberger-de Groot, A. C., Sariola, H., Friedman, R., Boivin, G. P., Cardell, E. L., and Doetschman, T. (1997). TGF-β2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes. Development *124*, 2659-2670.

Schmierer, B., and Hill, C. S. (2005). Kinetic analysis of Smad nucleocytoplasmic shuttling reveals a mechanism for transforming growth factor β-dependent nuclear accumulation of Smads. Mol Cell Biol *25*, 9845-9858.

Sehgal, I., and Thompson, T. C. (1999). Novel regulation of type IV collagenase (matrix metalloproteinase-9 and -2) activities by transforming growth factor-β1 in human prostate cancer cell lines. Mol Biol Cell *10*, 407-416.

Seoane, J., Le, H. V., Shen, L., Anderson, S. A., and Massague, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell *117*, 211-223.

Seoane, J., Pouponnot, C., Staller, P., Schader, M., Eilers, M., and Massague, J. (2001). TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. Nat Cell Biol *3*, 400-408.

Shah, M., Foreman, D. M., and Ferguson, M. W. (1994). Neutralising antibody to TGF-β1/2 reduces cutaneous scarring in adult rodents. J Cell Sci *107 (Pt 5)*, 1137-1157.

Shen, X., Hu, P. P., Liberati, N. T., Datto, M. B., Frederick, J. P., and Wang, X. F. (1998). TGF-β-induced phosphorylation of Smad3 regulates its interaction with coactivator p300/CREB-binding protein. Mol Biol Cell *9*, 3309-3319.

Shen, X., Li, J., Hu, P. P., Waddell, D., Zhang, J., and Wang, X. F. (2001). The activity of guanine exchange factor NET1 is essential for transforming growth factor-β-mediated stress fiber formation. J Biol Chem *276*, 15362-15368.

Shi, W., Sun, C., He, B., Xiong, W., Shi, X., Yao, D., and Cao, X. (2004). GADD34-PP1c recruited by Smad7 dephosphorylates TGF-β type I receptor. J Cell Biol *164*, 291-300.

Shi, Y., and Massague, J. (2003). Mechanisms of TGF-β signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N. P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-β signaling. Cell *94*, 585-594.

Shioda, T., Lechleider, R. J., Dunwoodie, S. L., Li, H., Yahata, T., de Caestecker, M. P., Fenner, M. H., Roberts, A. B., and Isselbacher, K. J. (1998). Transcriptional activating activity of Smad4: roles of SMAD hetero-oligomerization and enhancement by an associating transactivator. Proc Natl Acad Sci U S A *95*, 9785-9790.

Siegel, P. M., and Massague, J. (2003). Cytostatic and apoptotic actions of TGF-β in homeostasis and cancer. Nat Rev Cancer *3*, 807-821.

Slack, J. (2001). Essential Developmental Biology, 1st ed. Blackwell Publishing.

Stamenkovic, I. (2000). Matrix metalloproteinases in tumor invasion and metastasis. Semin Cancer Biol *10*, 415-433.

Strewler, G.J. and Greenspan, F.S. (1997). Basic and Clinical Endocrinology, 5th ed. Lange.

Stroschein, S. L., Bonni, S., Wrana, J. L., and Luo, K. (2001). Smad3 recruits the anaphase-promoting complex for ubiquitination and degradation of SnoN. Genes Dev *15*, 2822-2836.

Stroschein, S. L., Wang, W., Zhou, S., Zhou, Q., and Luo, K. (1999). Negative feedback regulation of TGF-β signaling by the SnoN oncoprotein. Science *286*, 771-774.

Subramanian, G., Schwarz, R. E., Higgins, L., McEnroe, G., Chakravarty, S., Dugar, S., and Reiss, M. (2004). Targeting endogenous transforming growth factor beta receptor signaling in SMAD4-deficient human pancreatic carcinoma cells inhibits their invasive phenotype. Cancer Res *64*, 5200-5211.

Sun, P. D., and Davies, D. R. (1995). The cystine-knot growth-factor superfamily. Annu Rev Biophys Biomol Struct *24*, 269-291.

Sun, Y., Liu, X., Eaton, E. N., Lane, W. S., Lodish, H. F., and Weinberg, R. A. (1999). Interaction of the Ski oncoprotein with Smad3 regulates TGF-β signaling. Mol Cell *4*, 499-509.

Tachibana, I., Imoto, M., Adjei, P. N., Gores, G. J., Subramaniam, M., Spelsberg, T. C., and Urrutia, R. (1997). Overexpression of the TGF-β-regulated zinc finger encoding gene, TIEG, induces apoptosis in pancreatic epithelial cells. J Clin Invest *99*, 2365-2374.

Taipale, J., Saharinen, J., and Keski-Oja, J. (1998). Extracellular matrix-associated transforming growth factor-β: role in cancer cell growth and invasion. Adv Cancer Res *75*, 87-134.

Takekawa, M., Tatebayashi, K., Itoh, F., Adachi, M., Imai, K., and Saito, H. (2002). Smad-dependent GADD45beta expression mediates delayed activation of p38 MAP kinase by TGF-β. EMBO J *21*, 6473-6482.

Takeshita, A., Yen, P. M., Misiti, S., Cardona, G. R., Liu, Y., and Chin, W. W. (1996). Molecular cloning and properties of a full-length putative thyroid hormone receptor coactivator. Endocrinology *137*, 3594-3597.

ten Dijke, P., and Hill, C. S. (2004). New insights into TGF-β-Smad signalling. Trends Biochem Sci *29*, 265-273.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN-β gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Thiery, J. P. (2003). Epithelial-mesenchymal transitions in development and pathologies. Curr Opin Cell Biol *15*, 740-746.

Tolwinski, N. S., and Wieschaus, E. (2004). Rethinking WNT signaling. Trends Genet *20*, 177-181.

Tomita, M., Choe, J., Tsukazaki, T., and Mori, N. (2004). The Kaposi's sarcoma-associated herpesvirus K-bZIP protein represses transforming growth factor beta signaling through interaction with CREB-binding protein. Oncogene *23*, 8272-8281.

Topper, J. N., DiChiara, M. R., Brown, J. D., Williams, A. J., Falb, D., Collins, T., and Gimbrone, M. A., Jr. (1998). CREB binding protein is a required coactivator for Smad-dependent, transforming growth factor beta transcriptional responses in endothelial cells [published erratum appears in Ptov Natl Acad Sci U S A 1998 Oct 13;95(21):12735]. Proc Natl Acad Sci U S A *95*, 9506-9511.

Trindade, M., Messenger, N., Papin, C., Grimmer, D., Fairclough, L., Tada, M., and Smith, J. C. (2003). Regulation of apoptosis in theXenopus embryo by Bix3. Development *130*, 4611-4622.

Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L., and Wrana, J. L. (1998). SARA, a FYVE domain protein that recruits Smad2 to the TGFβ receptor. Cell *95*, 779-791.

Uchida, K., Suzuki, H., Ohashi, T., Nitta, K., Yumura, W., and Nihei, H. (2001). Involvement of MAP kinase cascades in Smad7 transcriptional regulation. Biochem Biophys Res Commun *289*, 376-381.

Uesugi, M., Nyanguile, O., Lu, H., Levine, A. J., and Verdine, G. L. (1997). Induced alpha helix in the VP16 activation domain upon binding to a human TAF. Science *277*, 1310-1313.

Ulloa, L., Doody, J., and Massague, J. (1999). Inhibition of transforming growth factor-β/SMAD signalling by the interferon-gamma/STAT pathway. Nature *397*, 710-713.

Valderrama-Carvajal, H., Cocolakis, E., Lacerte, A., Lee, E. H., Krystal, G., Ali, S., and Lebrun, J. J. (2002). Activin/TGF-β induce apoptosis through Smad-dependent expression of the lipid phosphatase SHIP. Nat Cell Biol *4*, 963-969.

Vega, R. B., Matsuda, K., Oh, J., Barbosa, A. C., Yang, X., Meadows, E., McAnally, J., Pomajzl, C., Shelton, J. M., Richardson, J. A., *et al.* (2004). Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. Cell *119*, 555-566.

Ventura, J. J., Kennedy, N. J., Flavell, R. A., and Davis, R. J. (2004). JNK regulates autocrine expression of TGF-β1. Mol Cell *15*, 269-278.

Verrecchia, F., Tacheau, C., Schorpp-Kistner, M., Angel, P., and Mauviel, A. (2001a). Induction of the AP-1 members c-Jun and JunB by TGF-β/Smad suppresses early Smad-driven gene activation. Oncogene *20*, 2205-2211.

Verrecchia, F., Vindevoghel, L., Lechleider, R. J., Uitto, J., Roberts, A. B., and Mauviel, A. (2001b). Smad3/AP-1 interactions control transcriptional responses to TGF-β in a promoter-specific manner. Oncogene *20*, 3332-3340.

Vinals, F., and Pouyssegur, J. (2001). Transforming growth factor β1 (TGF-β1) promotes endothelial cell survival during in vitro angiogenesis via an autocrine mechanism implicating TGF-alpha signaling. Mol Cell Biol *21*, 7218-7230.

Waldrip, W. R., Bikoff, E. K., Hoodless, P. A., Wrana, J. L., and Robertson, E. J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. Cell *92*, 797-808.

Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kalpana, G. V., Goff, S. P., Yaniv, M., *et al.* (1996). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. EMBO J *15*, 5370-5382.

Wang, W., Mariani, F. V., Harland, R. M., and Luo, K. (2000). Ski represses bone morphogenic protein signaling in Xenopus and mammalian cells. Proc Natl Acad Sci U S A *97*, 14394-14399.

Warner, B. J., Blain, S. W., Seoane, J., and Massague, J. (1999). Myc downregulation by transforming growth factor β required for activation of the p15(Ink4b) G(1) arrest pathway. Mol Cell Biol *19*, 5913-5922.

Watabe, T., Kim, S., Candia, A., Rothbacher, U., Hashimoto, C., Inoue, K., and Cho, K. W. (1995). Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between *Xenopus* and mouse. Genes Dev *9*, 3038-3050.

Watanabe, M., Masuyama, N., Fukuda, M., and Nishida, E. (2000). Regulation of intracellular dynamics of Smad4 by its leucine-rich nuclear export signal. EMBO Rep *1*, 176-182.

Watanabe, M., and Whitman, M. (1999). FAST-1 is a key maternal effector of mesoderm inducers in the early Xenopus embryo. Development *126*, 5621-5634.

Weis-Garcia, F., and Massagué, J. (1996). Complementation between kinase-defective and activation defective TGF-β receptors reveals a novel form of receptor cooperativity essential for signaling. EMBO J *15*, 276-289.

Whitman, M. (2001). Nodal signaling in early vertebrate embryos: themes and variations. Dev Cell *1*, 605-617.

Wick, W., Grimmel, C., Wild-Bode, C., Platten, M., Arpin, M., and Weller, M. (2001). Ezrin-dependent promotion of glioma cell clonogenicity, motility, and invasion mediated by BCL-2 and transforming growth factor-β2. J Neurosci *21*, 3360-3368.

Wicks, S. J., Lui, S., Abdel-Wahab, N., Mason, R. M., and Chantry, A. (2000). Inactivation of smad-transforming growth factor beta signaling by Ca^{2+}-calmodulin-dependent protein kinase II. Mol Cell Biol *20*, 8103-8111.

Wilkinson, D. S., Ogden, S. K., Stratton, S. A., Piechan, J. L., Nguyen, T. T., Smulian, G. A., and Barton, M. C. (2005). A direct intersection between p53 and transforming growth factor beta pathways targets chromatin modification and transcription repression of the alpha-fetoprotein gene. Mol Cell Biol *25*, 1200-1212.

Williams, R. L., and Katan, M. (1996). Structural views of phosphoinositide-specific phospholipase C: signalling the way ahead. Structure *4*, 1387-1394.

Wisotzkey, R. G., Mehra, A., Sutherland, D. J., Dobens, L. L., Liu, X., Dohrmann, C., Attisano, L., and Raftery, L. A. (1998). Medea is a Drosophila Smad4 homolog that is differentially required to potentiate DPP responses. Development *125*, 1433-1445.

Wong, C., Rougier-Chapman, E. M., Frederick, J. P., Datto, M. B., Liberati, N. T., Li, J. M., and Wang, X. F. (1999). Smad3-Smad4 and AP-1 complexes synergize in transcriptional activation of the c-Jun promoter by transforming growth factor β. Mol Cell Biol *19*, 1821-1830.

Wotton, D., Lo, R. S., Lee, S., and Massagué, J. (1999a). A Smad transcriptional corepressor. Cell *97*, 29-39.

Wotton, D., Lo, R. S., Swaby, L. A., and Massague, J. (1999b). Multiple modes of repression by the Smad transcriptional corepressor TGIF. J Biol Chem *274*, 37105-37110.

Woychik, N. A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453-463.

Wu, G., Chen, Y. G., Ozdamar, B., Gyuricza, C. A., Chong, P. A., Wrana, J. L., Massague, J., and Shi, Y. (2000). Structural basis of Smad2 recognition by the Smad anchor for receptor activation. Science *287*, 92-97.

Wu, J. W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., Kyin, S., Muir, T. W., Fairman, R., *et al.* (2001). Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF- β signaling. Mol Cell *8*, 1277-1289.

Wu, J. W., Krawitz, A. R., Chai, J., Li, W., Zhang, F., Luo, K., and Shi, Y. (2002). Structural mechanism of Smad4 recognition by the nuclear oncoprotein Ski: insights on Ski-mediated repression of TGF-β signaling. Cell *111*, 357-367.

Xiao, Z., Latek, R., and Lodish, H. F. (2003). An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity. Oncogene *22*, 1057-1069.

Xiao, Z., Liu, X., Henis, Y. I., and Lodish, H. F. (2000a). A distinct nuclear localization signal in the N terminus of Smad 3 determines its ligand-induced nuclear translocation. Proc Natl Acad Sci U S A *97*, 7853-7858.

Xiao, Z., Liu, X., and Lodish, H. F. (2000b). Importin β mediates nuclear translocation of Smad3. J Biol Chem *275*, 23425-23428.

Xu, L., Alarcon, C., Col, S., and Massague, J. (2003). Distinct domain utilization by Smad3 and Smad4 for nucleoporin interaction and nuclear import. J Biol Chem *278*, 42569-42577.

Xu, L., Chen, Y. G., and Massague, J. (2000). The nuclear import function of Smad2 is masked by SARA and unmasked by TGFβ-dependent phosphorylation. Nat Cell Biol *2*, 559-562.

Xu, L., Kang, Y., Col, S., and Massague, J. (2002). Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGF-β signaling complexes in the cytoplasm and nucleus. Mol Cell *10*, 271-282.

Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L., and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. Genes Dev *12*, 2354-2370.

Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M., and Miyazono, K. (1999). Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. J Biol Chem *274*, 703-709.

Yahata, T., de Caestecker, M. P., Lechleider, R. J., Andriole, S., Roberts, A. B., Isselbacher, K. J., and Shioda, T. (2000). The MSG1 non-DNA-binding transactivator binds to the p300/CBP coactivators, enhancing their functional link to the Smad transcription factors. J Biol Chem *275*, 8825-8834.

Yamaguchi, K., Nagai, S., Ninomiya-Tsuji, J., Nishita, M., Tamai, K., Irie, K., Ueno, N., Nishida, E., Shibuya, H., and Matsumoto, K. (1999). XIAP, a cellular member of the inhibitor of apoptosis protein family, links the receptors to TAB1-TAK1 in the BMP signaling pathway. EMBO J *18*, 179-187.

Yamaguchi, K., Shirakabe, K., Shibuya, H., Irie, K., Oishi, I., Ueno, N., Taniguchi, T., Nishida, E., and Matsumoto, K. (1995). Identification of a member of the MAPKKK family as a potential mediator of the TGF-β signal transduction. Science *270*, 2008-2011.

Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y., and Hamada, H. (2001). The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. Genes Dev *15*, 1242-1256.

Yamashita, H., ten Dijke, P., Franzen, P., Miyazono, K., and Heldin, C. H. (1994). Formation of hetero-oligomeric complexes of type I and type II receptors for transforming growth factor-β. J Biol Chem *269*, 20172-20178.

Yang, X. J., Ogryzko, V. V., Nishikawa, J., Howard, B. H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319-324.

Yingling, J. M., Datto, M. B., Wong, C., Frederick, J. P., Liberati, N. T., and Wang, X. F. (1997). Tumor suppressor Smad4 is a transforming growth factor beta-inducible DNA binding protein. Mol Cell Biol *17*, 7019-7028.

Yu, Q., and Stamenkovic, I. (2000). Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-β and promotes tumor invasion and angiogenesis. Genes Dev *14*, 163-176.

Yue, J., and Mulder, K. M. (2000). Activation of the mitogen-activated protein kinase pathway by transforming growth factor-β. Methods Mol Biol *142*, 125-131.

Zacchigna, L., Vecchione, C., Notte, A., Cordenonsi, M., Dupont, S., Maretto, S., Cifelli, G., Ferrari, A., Maffei, A., Fabbro, C., *et al.* (2006). Emilin1 links TGF-β maturation to blood pressure homeostasis. Cell *124*, 929-942.

Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B., and Kern, S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell *1*, 611-617.

Zhang, X., Krutchinsky, A., Fukuda, A., Chen, W., Yamamura, S., Chait, B. T., and Roeder, R. G. (2005). MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. Mol Cell *19*, 89-100.

Zhang, Y., Feng, X. H., and Derynck, R. (1998). Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-β-induced transcription. Nature *394*, 909-913.

Zhou, S., Zawel, L., Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Characterization of human FAST-1, a TGF-β and activin signal transducer. Mol Cell *2*, 121-127.

References

Zwilling, S., Annweiler, A., and Wirth, T. (1994). The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP. Nucleic Acids Res *22*, 1655-1662.